Document Title:	**Validation of a Confirmatory Proteomic Mass Spectrometry Body Fluid Assay for Use in Publicly Funded Forensic Laboratories**
Author(s):	**Donald Siegel, Ph.D.**
Document Number:	**310765**
Date Received:	**September 2025**
Award Number:	**15PNIJ-21-GG-02712-SLFO**

Research Final Report

Federal Agency and Organization:	**U.S. Department of Justice, Office of Justice Programs,** **National Institute of Justice**
Federal Grant Name:	**NIJ FY 21 Research and Evaluation for the Testing and Interpretation of Physical Evidence in Publicly Funded Forensic Science Laboratories**
Federal Grant Number:	**15PNIJ-21-GG-02712-SLFO**
Project Title:	**Validation of a Confirmatory Proteomic Mass Spectrometry Body Fluid Assay for Use in Publicly Funded Forensic Laboratories**
Principal Investigator:	**Donald Siegel, Ph.D., City Research Scientist** **Department of Forensic Biology** **Email: DSiegel@ocme.nyc.gov**
Submission Date:	**June 9, 2025**
DUNS:	**806056446**
EIN:	**13-6400434**
Recipient Organization:	**City of New York, Office of Chief Medical Examiner**
Project/Grant Period:	**01/01/2021 to 12/31/2024**
Reporting Period End Date:	**12/31/2024**
Award Amount	**$105,592**

Abbreviations

ABC = ammonium bicarbonate
ANAB = ANSI National Accreditation Board
ANOVA = analysis of variance
BCA = bicinchoninic acid
BSA = bovine serum albumin
CV = coefficient of variance
FWHM = full width at half maximum
HPLC = high performance liquid chromatography
LOB = limit of blank
LOD = limit of detection
LOQ = limit of quantitation
MRM = multiple reactions monitoring
MS/MS = tandem mass spectrometry
NR = non-redundant
NYC OCME = New York City Office of Chief Medical Examiner
PCM = peptide calibration mixture
SDC = sodium deoxycholate

Final Report - This Final Report follows NIJ's guidelines for Post-Award Requirements for Research, Development, and Evaluation Grants for the Final Research Report and is divided into five sections, **1**) Summary of the Project, **2**) Participants and Other Collaborating Organizations, **3**) Changes in Approach from Original Design and Reason for Change (if applicable), **4**) Outcomes, and **5**) Artifacts, **6**) Supplementary Data. Sections are subdivided as described in the guidelines.

1) Summary of the Project – This section is divided into four subsections as described in the guidelines: **1.1**) Major goals and objectives, **1.2**) Research questions, **1.3**) Research design, methods, analytical and data analysis techniques, and **1.4**) Expected applicability of the research.

1.1. MAJOR GOALS AND OBJECTIVES – The overarching objective of this grant was to validate a confirmatory assay for body fluid identification that is both scientifically and statistically sound and can be implemented in public forensic laboratories. These goals have been achieved (discussed in subsection 1.4. Applicability of the Research), as the assay has been accredited by the national forensic accrediting body ANSI National Accreditation Board (**ANAB**), approved by the New York State Commission on Forensic Science, and implemented into casework here at the New York City Office of Chief Medical Examiner (**NYC OCME**). Additionally, with respect to dissemination, our validation and protocols have been made available to other forensic laboratories through the NYC OCME public facing website – links below.

https://www.nyc.gov/assets/ocme/downloads/pdf/molecular_serology_validation.pdf,

https://www.nyc.gov/site/ocme/services/fbio-2025-molecular-serology.page

BACKGROUND: The three body fluids most commonly examined by forensic laboratories are blood, saliva and semen. Except for microscopic identification of sperm, currently there are no routinely used confirmatory tests for these body fluids. Consequently, while DNA from evidence collected at a crime scene may be able to identify an individual, the source of that individual's DNA cannot

be stated conclusively – i.e. was it from blood, saliva or semen, or shed epithelial cells or secondary DNA transfer? This is a particular problem now as DNA testing is becoming increasingly more sensitive – able to identify an individual from a small number of their cells. Consequently, a confirmatory test for blood, saliva and semen could be valuable in the fair administration of justice – e.g. by aiding law enforcement early in investigations, offering context to what may have occurred during a purported crime, offering scientifically and statistically sound evidence in court, and aiding the unjustly accused.

Multiple aspects of an assay must come together before it can be introduced into routine testing in a forensic laboratory. Not only must specific body fluid markers be identified and evaluated and a method chosen to analyze them (this was done with previous NIJ basic research grants 2014-BX-K014 and 2017-NE-BX-0003), but all features of the assay must be evaluated not only from sample extraction, processing and analysis, but in establishing the quality metrics of how samples are consistently and reliably analyzed, as well as the metrics for evaluating and monitoring instrumentation and software. The specific aims of this application focused on the quality control aspects of the assay that need to be put in place for its successful introduction into routine casework.

1.2 RESEARCH QUESTIONS – Body fluid detection in our assay system is based on the identification of proteins that have specific functions within each body fluid and are therefore highly enriched in their respective body fluids – e.g. hemoglobin a protein in blood carries oxygen throughout the body, amylase a protein in saliva begins the digestion process, and semenogelin a protein in semen is required for reproduction. To ensure accurate body fluid identification, multiple specific proteins are evaluated in each body fluid. A list of all the marker proteins used in this assay to identify blood, saliva and semen may be found in Table 1.

Table 1 - Body Fluid Marker & Standards Peptides with Transition Ions

Body Fluid/Standard	Protein Marker	Accession No.	Peptide Sequence	Transitions
Blood	Hemoglobin subunit beta (**HBB**)	P68871	SAVTALWGK	446.764/774.4509 446.764/675.3824 446.764/574.3348 446.764/390.2136
			VNVDEVGGEALGR	657.836/887.4581 657.836/758.4155 657.836/659.3471 657.836/416.2616
			LLVVYPWTQR	637.866/949.489 637.866/850.4206 637.866/687.3573 637.866/326.2438
	Hemoglobin subunit alpha (**HBA**)	P69905	VGAHAGEYGAEALER	510.583/745.3839 510.583/617.3253 510.583/488.2827 510.583/842.3791
			TYFPHFDLSHGSAQVK	611.969/589.3304 611.969/711.8599 611.969/663.3335 611.969/785.3941
			FLASVSTVLTSK	626.861/992.5623 626.861/921.5251 626.861/834.4931 626.861/735.4247
	Band 3 anion transport protein (**SLC4A1**)	P02730	IPPDSEATLVLVGR	733.914/828.5302 733.914/677.3723 733.914/628.8459 733.914/757.4931
			ADFLEQPVLGFVR	745.904/915.5411 745.904/787.4825 745.904/591.3613 745.904/478.2772
			ASTPGAAAQIQEVK	685.867/957.5364 685.867/886.4993 685.867/815.4621 685.867/556.3089
	Albumin (**ALBU**)	P02768	LVNEVTEFAK	575.311/937.4625 575.311/823.4196 575.311/694.3770 575.311/595.3086
			AEFAEVSK	440.7242/680.3614 440.7242/533.2930 440.7242/462.2558 440.7242/348.1554

Saliva	Alpha-amylase 1 (**AMY1A**)	P04745	LSGLLDLALGK	550.34/986.5881 550.34/616.3665 550.34/729.4505 550.34/899.5560
			ALVFVDNHDNQR	476.572/449.6917 476.572/622.2944
				476.572/572.7601 476.572/499.2259
			IYVSDDGK	448.722/434.1882 448.722/620.2886 448.722/521.2202 448.722/376.2231
	Cystatin SA (**CYS2**)	P09228	ALHFVISEYNK	440.903/640.2937 440.903/568.3242 440.903/322.1874 440.903/469.2558
			ATEDEYYR	523.725/874.3577 523.725/745.3151 523.725/630.2882 523.725/501.2456
			SQPNLDTC[CAM]AFHEQPELQK	714.669/614.3508 714.669/963.9544 714.669/858.4065 714.699/801.8645
	Histatin 1 (**HTN1**)	P15515	EFPFYGDYGSNYLYDN	982.405/524.2350 982.405/411.1510 982.405/248.0880 982.405/277.1180
	Protein LEG1 homolog (**C6orf58**)	Q6P5S2	ESPGQLSDYR	576.27/935.4581 576.27/838.4054 576.27/338.1823 576.27/653.3253
Standards	Cytochrome C (**CYC**)	P62894	TGQAPGFSYTDANK	728.8388/798.3628 728.8388/585.7749 728.8388/550.2564 728.8388/358.1721

Semen	Semenogelin 2 (**SEMG2**)	Q02383	GQLPSGSSQFPHGQK	777.887/713.3729 777.887/566.3045 777.887/685.3466 777.887/628.8046
			LWVHGLSK	470.274/640.3777 470.274/541.3093 470.274/404.2504 470.274/826.4570
			GSISIQTEEK	546.283/834.4203 546.283/747.3883 546.283/634.3042 546.283/506.2457
	Semenogelin 1 (**SEMG1**)	P04279	HLGGSQQLLHNK	444.577/511.2987 444.577/541.2911 444.577/624.3828 444.577/708.3424
			SQIQAPNPK	491.769/767.4410 491.769/654.3570 491.769/526.2984 491.769/455.2613
			EQDLLSHEQK	613.804/969.5000 613.804/741.3890 613.804/628.3049 613.804/541.2729
	Prostate specific antigen (**KLK3**)	P07288	IVGGWEC	539.255/964.4193 539.255/865.3509 539.255/808.3294 539.255/436.1860
			SVILLGR	379.25/571.3926 379.25/458.3085 379.25/345.2245 379.25/300.1918
			LSEPAELTDAVK	636.838/943.5095 636.838/775.4196 636.838/646.3770 636.838/533.2930

The questions posed in this application focused on four areas of the assay's performance each of which is discussed in detail in section 1.3 below. Overall, these included **1)** evaluation of general assay performance metrics of marker detection - e.g. what is the assay's sensitivity,

accuracy, precision, etc., how does the assay perform with different types of samples – e.g. neat, mock, and aged samples, as well as how it performs with non-targeted body fluids (e.g. menstrual blood and vaginal fluid) or body fluids from other species, **2)** because the assay tests forensic samples which may be exposed to unknown proteins in the environment, what is the likelihood that one of the body fluid marker peptides may be present from another organism in the environment, **3)** what is the potential false negative rate due to protein polymorphisms, and **4**) which assay quality metrics need to be monitored routinely and which for long term trends?

1.3 RESEARCH DESIGN, METHODS, ANALYTICAL AND DATA ANALYSIS TECHNIQUES – As described above, this section is divided into four pats: **1**) Evaluation of Assay Performance, **2**) Likelihood of a Non-Human Peptides Producing a False Positive, **3**) Evaluation of a False Negative Rate Due to Protein Polymorphisms, and **4**) Establishing Assay Quality Metrics Needed for Routine Monitoring.

1.3.1 Evaluation of General Assay Performance Metrics of Marker Detection – Sensitivity, Accuracy, Precision, Limits of Quantitation, Limits of Detection.

The body fluid assay is composed of two distinct parts. The first (**1.3.1.1**) is a quantitative assay to determine the amount of protein extracted from a sample. The second (**1.3.1.2**) is a classification assay to determine which body fluid (blood, saliva or semen) is present in the sample. These two assays, quantitative and classification, are evaluated by different methods, and consequently will be described separately below.

1.3.1.1 Evaluation of the Quantitative Assay – The bicinchoninic acid (**BCA**) assay is used to quantify proteins. The assay uses bovine serum albumin (**BSA**) as a standard protein against which body fluid proteins are compared to determine their protein concentrations. Consequently, validation of the BCA assay was done in two parts. Part 1 evaluated the BSA standard to determine the sensitivity, precision, and repeatability of the quantitation assay. Part 2 used blood, saliva, and semen samples to evaluate precision and repeatability of measuring sample proteins.

1.3.1.1.1 – BCA Protein Quantitation Assay Evaluating Standard Protein BSA

• SENSITIVITY – The BCA assay has a linear working range from 20 to 2,000 ng protein/µl. To validate sensitivity, BSA standard was serially diluted from 2,000 ng/µl to 10 ng/µl (i.e., below the manufacturer's stated limit of detection). Samples were run in triplicate on 6 different days by two analysts (468 total samples). These data were used to establish the linear range, limit of quantitation (**LOQ**), and limit of detection of the assay (**LOD**). Measurements of 0 µg/ml of BSA (water and extraction solution (1% sodium deoxycholate in 50 mM ammonium bicarbonate **(ABC/SDC**)) were used to establish the highest expected absorbance in the absence of protein in a sample (limit of blank, **LOB**). Variation at low concentrations of BSA standards was used to establish LOD above LOB.

The LOD for BSA was determined to be $\geq$ an absorbance of 0.136, which is equivalent to ~50 µg/ml of BSA. Based on the low coefficient of variance (**CV**) for 50 µg/ml of BSA, the LOQ is functionally the same as LOD at 50 µg/ml of BSA.

• REPEATABILITY & PRECISION – Repeatability (a subset of precision, measuring variation in measurements obtained by the same method in a short time frame) was evaluated by analysis of variance (**ANOVA**) using the BSA sensitivity validation data generated above. To evaluate the repeatability of BSA standard and BCA plate preparation, only measurements of BSA batches that were run on BCA plates on three days were used.

The CVs between batch variation and between plate variation were low, indicating that these procedures do not introduce unacceptable variation to the measurement. Total variation is within tolerance. Considering repeat measurements of the same preparation of BSA on the sample plate, variation is higher at the lowest BSA concentrations. However, only 0 µg/mL had a CV above 15%. Linearity, the R^2 value for all standard measurements across multiple BSA

preparations and plate runs is 0.989. The mean R^2 value per BSA batch and plate, which are the conditions in which the standards curve for sample evaluation is established, was 0.993.

1.3.1.1.2 – BCA Protein Quantitation Assay Evaluating Body Fluids - Sensitivity, precision, and repeatability of the BCA assay were measured using neat samples of blood, saliva, and semen. Samples were extracted following NYC OCME Proteomic Body Fluid SOPs. All analyses were performed on supernatants of extracted samples.

• **SENSITIVITY** – The amount of protein in blood, saliva, and semen varies on a volume basis with neat blood having ~250 ng protein/µl, semen ~25 ng protein/µl, and saliva ~1.5 ng protein/µl. The protein concentration of saliva is significantly more variable than semen or blood. Consequently, two donor samples were used to assay protein content in blood and semen while three donor samples were used for measurements in saliva. Based on the LOD established in section 1.3.1.1.1, measurements of semen, saliva, and blood were assigned as detectable as defined by an absorbance greater than 0.14 with a predicted concentration above 50 µg/ml. Multiple dilutions of each body fluid were prepared within the BCA assay's working range as well as near its LOQ as established using BSA in Section 1.3.1.1.1. Each dilution of each sample was performed in triplicate on six different days by two analysts. Samples were analyzed and used to establish linearity and approximate (since exact body fluid protein content is unknown) LOQ and LOD for all body fluids. As expected, due to the inherent differences in protein concentrations between body fluids, neat blood could be diluted at least forty times (1+39) and still be detected. Neat semen could be diluted twice (1+1) and still be detected, while neat saliva was below the level of detection of the BCA assay (see Supplemental Table 1).

• **PRECISION AND REPEATABILITY** – Precision and repeatability analyses were done on all samples described above. For samples where the same extraction and dilution were measured on the same plate at least 3 times, CVs were calculated for plate, extraction, and donor. For blood and

semen, absorbance CVs were higher for different extractions from the same donor, but consistently under 10%. For saliva, absorbance CVs are low for repeat measurements of the same extracion, but variation increases more dramatically for different extractions from the same donor (see Supplemental Tables 2-4).

1.3.1.2 Body Fluid Classification Assays – This section is divided into eight parts: 1) Definition of Classification Assays and the Metrics Used to Evaluate Them, 2) Marker Identification, 3) Neat Samples, 4) Mock Samples, 5) Mixed Samples, 6) Non-Targeted Samples, 7) Aged Samples, and 8) Data Analysis, 9) Body Fluid Identification Performance Metrics.

1.3.1.2.1 Classification Assay and the Metrics - A classification assay determines which body fluid (blood, saliva or semen) is present in the sample. Classification assays are not quantitative. They are based on identifying a sample result as a true positive (positive sample with positive test result), a false positive (negative sample with positive result), a true negative (negative sample with negative result), or a false negative (positive sample with negative result). For each body fluid in this validation, positive samples are all known samples of that body fluid. Negative samples are the two other body fluids plus negative controls and non-target samples (e.g., vaginal fluid and non-human blood, saliva, and semen). True positives and true negatives are correct results, false positives and false negatives are errors. Classification metrics based on these results are ratios (often reported as percentages), and defined as follows:

- **ACCURACY:** The total correct classifications / total samples tested
- **SENSITIVITY:** Sensitivity is total true positives / total positive samples tested (true positives and false negatives) and is the inverse of the false negative rate.
- **FALSE NEGATIVE RATE:** The false negative rate is total false negatives / total positive samples tested and is the inverse of sensitivity.

- **SPECIFICITY:** Specificity is total true negative / all negative samples tested (true negatives and false positives) and is the inverse of false positive rate.
- **FALSE POSITIVE RATE:** The false positive rate is the is the total false positives / all negative samples tested and is the inverse of specificity.
- **FALSE DISCOVERY RATE:** The false discovery rate is the total false positives / all positive results obtained, and the primary metric used in this validation to understand false positives (or in this case lack thereof) and the confidence that can be applied to that result.
- **REPEATABILITY:** Repeatability in classification assays is the ability of the assay to obtain consistent classification results for replicate measurements of the same sample. Repeatability of the assay will be assessed at the level of analyst, extraction batch, and HPLC column injection using the samples described above.

1.3.1.2.2 Marker Identification - Following protein extraction and quantitation, samples were processed (reduced and alkylated) and then digested with trypsin (using OCME SOPs). Tryptic peptides were then separated by HPLC and analyzed by mass spectrometry. Body fluid identification is based on detection of body fluid specific peptide markers and their fragment ions (Table 1). A positive body fluid identification is made when all marker peptides and all peptide transition ions are detected. Validation of body fluid marker identification was performed on 72 body fluid samples, 20 of blood, 24 of saliva, and 27 of semen. Accuracy, sensitivity, specificity, and repeatability were assessed.

1.3.1.2.3 Neat Body Fluids: Neat samples from five individuals for each body fluid (blood, saliva, and semen) were extracted three separate times by two analysts following Proteomic Body Fluid SOPs. The remaining 15 samples of each blood, saliva, and semen were extracted once each by two analysts. An extraction negative control was included in each extraction batch. Sample supernatants from each extraction were analyzed from protein quantitation through HPLC-MS.

Each sample and negative control were spiked with a cytochrome C peptide (to measure HPLC sample loading). Each sample and negative control were run twice by HPLC-MS, once through each of the two HPLC's columns.

1.3.1.2.4 Mock body fluid samples consisted of DNA proficiency tests samples received by the Department of Forensic Biology between the years 2009 to 2020. It is important to note that while the type of body fluid used in preparing these samples was known, the amount of protein they contained was not. Additionally, these kits only provided samples for semen and blood. Saliva mock samples were created by collecting buccal swabs from volunteers. Known and non-probative/mock evidence samples of blood, saliva, and semen were also prepared by pipetting 20 µg, 10 µg, 2 µg, 1 µg, 0.5 µg, 0.25 µg, 0.1 µg, and 0.05 µg of protein (measured from neat body fluid samples) on to cotton swabs. Mock samples from 5 donors for each body fluid were extracted 3 times each according to Proteomic Body Fluid SOPs and run by two analysts. Each sample was run twice through HPLC-MS (once through each HPLC column.

1.3.1.2.5 Mixed Samples - In order to assess the classification assay's performance in the case of a sample containing a mixture of body fluids, mixtures of blood/semen, blood/saliva, and semen/saliva were prepared. Three samples from three different individuals of each blood, semen, and saliva were mixed in a 1:1 protein ratio, extracted, processed, digested, and prepared for HPLC-MS according to Proteomic Body Fluid SOPs. In addition, one sample from one individual was used to make mixtures of blood/semen, blood/saliva, and semen/saliva in 1:4, 1:8, 1:12, 12:1, 8:1, and 4:1 protein ratios, and blood/saliva/semen mixtures in 1:1:1, 1:1:10, 1:10:1, and 10:1:1 protein ratios. Each mixture was run twice through HLPC-MS, once through each column. Classification results for blood, semen, and saliva were analyzed as described above for neat samples.

1.3.1.2.6 Non-Targeted Samples - Non-targeted samples were human body fluids other than blood, semen, and saliva, as well as non-human body fluids. Human samples tested were vaginal fluid (expected negative for blood, saliva, and semen), and menstrual blood (expected positive for blood). Non-human samples were blood from Canis lupus (dog), Felis catus (cat), Bos taurus (cow), Sus scrofa (pig), and Gallus gallus (chicken), saliva from Canis lupus (dog) and Felis catus (cat), and semen from Sus scrofa (boar) and Bos taurus (oxen). All samples were extracted and analyzed by HPLC-MS according to Proteomic Body Fluid SOPs.

1.3.1.2.7 Aged Samples - Aged samples were OCME DNA proficiency test kits from Bode Technologies and Collaborative Testing Services, Inc. prepared from whole body fluids on cotton swabs, cards and fabric dating back to 2009. Aged mock samples were extracted according to Proteomic Body Fluid SOPs and run by two analysts. Each sample was run twice through HPLC-MS, once through each column.

1.3.1.2.8 Data Analysis: This assay functions as three binary classification tests, one each for blood, semen, and saliva (it is possible to have positive result for more than one classification in the case of a mixed sample). The performance metrics reported below are from results for neat, known, and non-probative/mock evidence samples (aged and non-aged), non-target samples, and mixed samples as described above.

The specific cutpoints for all peptide metrics (i.e. the threshold that must be met for each metric to determine that the peptide is present in the sample) were determined by values that optimized precision and sensitivity on multiple subsets of metric data for each peptide and averaging the results. The resampling and averaging process prevents overfitting. Expected ion ratios for each target body fluid peptide were generated from 396 HPLC-MS runs of labeled synthesized peptides. The ion ratios were largely consistent over all runs, with narrow distributions. The median over all runs was selected as the expected ratio for each fragment ion

which minimizes the effects of outliers. Decision criteria for identification of a peptide using the cutpoints, and cutpoints themselves, are found in the LCMS MRM Interpretation SOP.

1.3.1.2.9 Body Fluid Identification Performance Metrics **-** The validation includes a total of 1496 LCMS runs, each with associated preceding cleaning solution run. Of these, 3 runs were negative with no detection of Cytochrome C control, indicating failed or incomplete LCMS injection, and were removed from the data set. This results in a total of 1,493 LCMS runs of 691 sample extractions. Each body fluid (blood, saliva and semen) is discussed below.

- **Blood** – Table 2 below shows the performance of blood identification in 212 extractions of known human blood samples from 20 individuals, and 479 extractions of known non-blood or non-human blood samples. Non-blood samples include human saliva, semen, vaginal fluid, and negative controls. Non-human blood samples include those from livestock and domestic pets and one gorilla.

For a sample to be identified as positive for blood, that sample must have all nine blood peptide markers present, three from each blood marker protein.

Table 2 Performance Metrics of Neat Blood

Metric	%	95% CI Lower	95% CI Upper	N
Sensitivity	87.7%	82.4%	91.69%	212
False Discovery Rate	0.0%	0.0%	2.52%	186
Accuracy	96.2%	94.5%	97.48%	691
Injection Repeatability	100.0%	99.3%	100.00%	691

The sensitivity reported above includes all known blood samples, even those with very small known amounts of protein (to determine limits of detection, described above). Sensitivity is also reported for different types of known blood samples in Table 3.

Table 3 Performance Metrics of Varying Types of Blood Samples

Sample	%	95% CI Lower	95% CI Upper	N
Neat Single	100.0%	94.3%	100.0%	80
Neat Mixture	96.2%	78.4%	99.8%	26
Mock - Known Amount 1µg and less	20.0%	8.4%	39.1%	30
Mock - Known Amount >1µg	100.0%	85.9%	100.0%	30
Mock - Unknown Amount	97.8%	87.0%	99.9%	46

"N" indicates the denominator of the calculated performance metric. For sensitivity, N is the total number of known positive samples. For false discovery rate N is the total number samples with a positive result, and for accuracy N is the total samples tested. A 95% percent confidence interval is calculated for each performance metric using the Wilson score interval[1] with Yates continuity correction[2] for improved interval estimation for values near 0 and 100%.

- **Saliva -** Table 4 shows the performance of saliva identification in 170 extractions of known human saliva samples from 25 individuals, and 521 extractions of known non-saliva or non-human saliva samples. Non-saliva samples include blood, menstrual blood, semen, and vaginal fluid, and negative controls. Non-human saliva samples include those from domestic pets.

Table 4 Performance Metrics of Neat Saliva

Metric	%	95% CI Lower	95% CI Upper	N
Sensitivity	88.2%	82.2%	92.49%	170
False Discovery Rate	0.0%	0.0%	3.11%	150
Accuracy	97.1%	95.5%	98.18%	691
Injection Repeatability	96.7%	95.0%	97.83%	691

Due to variability of protein concentrations in saliva, a positive result for identification of saliva in a sample requires detection of a minimum of seven of the eight saliva peptide markers. All peptides

from amylase (AMY1) and cystatin (CYTT) proteins must be present (6 peptides, 3 per protein) and the 7th peptide may be either the histatin (HIS1) marker protein or the LEG1 (LEG1H) marker.

The sensitivity reported above includes all known neat saliva samples, even those with very small known amounts of protein (to determine limits of detection, described above). Sensitivity is therefore also reported below for different types of known saliva samples (Table 5).

Table 5 Performance Metrics of Varying Types of Saliva Samples

Sample	%	95% CI Lower	95% CI Upper	N
Neat Single	100.0%	93.9%	100.0%	74
Neat Mixture	100.0%	84.0%	100.0%	26
Mock - Known Amount 1µg and less	36.7%	20.5%	56.1%	30
Mock - Known Amount >1µg	96.7%	80.9%	99.8%	30
Mock - Unknown Amount	100.0%	65.5%	100.0%	10

"N" indicates the denominator of the calculated performance metric. For sensitivity, N is the total number of known positive samples. For false discovery rate N is the total number samples with a positive result, and for accuracy N is the total samples tested. A 95% percent confidence interval is calculated for each performance metric using the Wilson score interval[1] with Yates continuity correction[2] for improved interval estimation for values near 0 and 100%.

- **Semen –** Table 6 below shows the performance of semen identification in 196 extractions of known human semen samples from 27 individuals, and 495 extractions of known non-semen or non-human semen samples. Non-semen samples include saliva, blood, menstrual blood, and vaginal fluid, and negative controls. Non-human semen samples include those from livestock.

Due to variability of protein concentrations in semen, a positive result for identification of semen in a sample requires detection of a minimum of seven of the nine semen peptide markers. All three peptides from prostate specific antigen (KLK3) must be present and at least 2 out of 3 peptide markers from both semenogelin 1 and 2 (SEMG1 and SEMG2) must be identified.

Table 6 Performance Metrics of Neat Semen

Metric	%	95% CI Lower	95% CI Upper	N
Sensitivity	94.4%	89.9%	97.02%	196

Table 6 Performance Metrics of Neat Semen

Metric	%	95% CI Lower	95% CI Upper	N
False Discovery Rate	0.0%	0.0%	2.54%	185
Accuracy	98.4%	97.1%	99.16%	691
Injection Repeatability	99.0%	97.8%	99.56%	691

The sensitivity reported above includes all known semen samples, even those with very small known amounts of protein (to determine limits of detection, described above). Sensitivity is therefore also reported below for different types of known semen samples.

Table 7 Performance Metrics of Varying Types of Semen Samples

Sample	%	95% CI Lower	95% CI Upper	N
Neat Single	97.8%	91.4%	99.6%	90
Neat Mixture	100.0%	84.0%	100.0%	26
Mock - Known Amount 1µg and less	76.7%	57.3%	89.4%	30
Mock - Known Amount >1µg	100.0%	85.9%	100.0%	30
Mock - Unknown Amount	90.0%	66.9%	98.2%	20

"N" indicates the denominator of the calculated performance metric. For sensitivity, N is the total number of known positive samples. For false discovery rate N is the total number samples with a positive result, and for accuracy N is the total samples tested. A 95% percent confidence interval is calculated for each performance metric using the Wilson score interval[1] with Yates continuity correction[2] for improved interval estimation for values near 0 and 100%.

1.3.2 Likelihood of a Non-Human Peptides Producing a False Positive – Crime scenes can be open to the environmental contaminations from an unknowable number of biological samples. In order to evaluate the specificity of our human body fluid marker peptides with respect to the possibility that other peptides in nature might be similar to them, we performed a comparison of our marker peptide data to peptides in the biosphere. This was performed using Multiple Reaction Monitoring (**MRM**) transition data, which included parent and fragment ion masses, characteristic fragment ion intensity patterns, and expected retention times of our markers compared to a complete in silico trypsin digest of the NCBI non-redundant (**NR**) protein database (downloaded 5/31/2024 https://ftp.ncbi.nlm.nih.gov/) is necessary. The NR database contains 707,028,945 proteins

ranging in size from 386 to 98,182 amino acids, and 1,532,383 species. The in silico digestion generated all peptide sequences that would be produced by database proteins digested by trypsin, and was accomplished with Rapid Peptide Generator (https://rapid-peptide-generator.readthedocs.io/).

The resulting database of tryptic peptides was then filtered with an in-house Python-based filter program with the following steps: 1) Filter for peptide parent ion masses that correspond to the OCME's body fluid marker peptide masses (+-0.5 Da), 2) Generate all possible y and b fragment ions using the Python pepfrag library (y and b are fragments that include either the N- or C-terminus of the peptide and are the most common fragments generated by the mass spectrometry methods used by OCME Molecular Serology lab), 3) Filter possible fragments for those with masses (+/-0.5 Da) matching the four fragment ions per peptide targeted by the OCME Molecular Serology assay (see Molecular Serology Data Interpretation Manual), and 4) Filter for parents masses with fragments matching all four targeted fragment ions. This filtering process resulted in 127,747 candidate peptides.

Retention times of candidate peptides was predicted using Prosit[3] (https://koina.wilhelmlab.org/docs#post-/Prosit_2019_irt/infer), implemented in Python. Predicted retention times were calibrated to the 20-minute gradient used in the OCME Molecular Serology assay using the experimental retention time data from OCME body fluid marker peptides, and peptides were filtered for predicted, calibrated retention times +/- 1 minute of retention times of body fluid markers. After retention time filtering, the pool of candidate peptides was reduced to 14,938 sequences.

Candidate sequences were further evaluated by predicting fragment ion intensities using Prosit[3] (https://koina.wilhelmlab.org/docs#post-/Prosit_2019_intensity/infer). For each candidate, a set of four fragment ion intensities (spectrum) corresponding to targeted marker fragment ions were

compared to intensities of body fluid marker peptides by calculating the normalized contrast angle between the predicted candidate spectra and marker spectra, resulting in a score between 0-1, with 1 indicating identical spectra. Candidate peptides were then filtered for a comparison score of at least 0.6. The 0.6 threshold is lower than would typically be used in a comparison of two experimental spectra in order to account for the uncertainty inherent in spectral prediction. The resulting pool of 259 candidate peptides are therefore the peptides, out of all possible tryptic peptides in the NCBI protein database, that are most likely to mimic the signal of body fluid marker peptides, and as such are used to understand the probability of false positives for body fluid marker peptides. However, the true degree of similarity between any one candidate peptide and a body fluid marker can only be fully determined experimentally, and may not, in practice, be similar enough to the target spectra to cause a false positive.

Table 8 shows the number of occurrences of the marker peptide sequence, or sequences matching the characteristic mass spectrum of a marker in all known protein sequences other than the targeted body fluid proteins. The global probability upper bound is the upper bound of a Wilson confidence interval on the frequency of the signal in the database.

Table 8 Occurrence of Peptides with Spectra Corresponding to Body Fluid Marker Peptides

Protein	Peptide	Matching Seq. N*	Species N	Eukaryote Species	Bacteria Species	Virus/Archaea Species	Global Prob Upper Bound
B3AT	ADFLEQPVLGFVR	3	5	4	1	0	0.0000246
B3AT	ASTPGAAAQIQEVK	11	61	52	9	0	0.0001420
B3AT	IPPDSEATLVLVGR	6	99	94	5	0	0.0002157
HBB	LLVVYPWTQR	8	559	554	5	0	0.0010603
HBB	SAVTALWGK	7	35	32	3	0	0.0000899
HBB	VNVDEVGGEALGR	7	140	135	5	0	0.0002936
HBA	FLASVSTVLTSK	4	180	169	11	0	0.0003686
HBA	TYFPHFDLSHGSAQVK	5	160	157	3	0	0.0003312
HBA	VGAHAGEYGAEALER	13	58	51	7	0	0.0001361
AMY1	ALVFVDNHDNQR**	5	455	446	8	0	0.0008725
AMY1	IYVSDDGK**	42	92	26	64	1	0.0002023
AMY1	LSGLLDLALGK	12	33	18	15	0	0.0000858
CYTT	ALHFVISEYNK	1	1	0	1	0	0.0000140

Protein	Peptide	Matching Seq. N*	Species N	Eukaryote Species	Bacteria Species	Virus/Archaea Species	Global Prob Upper Bound
CYTT	ATEDEYYR**	11	38	22	16	0	0.0000960
HIS1	EFPFYGDYGSNYLYDN	1	3	3	0	0	0.0000195
LEG1H	ESPGQLSDYR	1	18	18	0	0	0.0000542
SEMG1	EQDLLSHEQK	1	23	23	0	0	0.0000649
SEMG1	HLGGSQQLLHNK	1	7	7	0	0	0.0000294
SEMG1	SQIQAPNPK	3	26	24	2	0	0.0000713
KLK3	IVGGWECEK	4	4	1	3	0	0.0000220
KLK3	LSEPAELTDAVK	12	25	18	7	0	0.0000692
KLK3	SVILLGR	93	1904	209	1645	34	0.0034504
SEMG2	GQLPSGSSQFPHGQK	1	6	6	0	0	0.0000270
SEMG2	GSISIQTEEK	4	25	19	6	0	0.0000692
SEMG2	LWVHGLSK	3	13	12	1	0	0.0000432
CYTT	SQPNLDTCAFHEQPELQK**	1	2	1	0	0	0.0000168

Note:

*Includes target sequences. Peptide with 'Matching Seq. N' = 1 matches only itself.

**Non-target human proteins each added as a species count for frequency calculation.

1.3.3 Evaluation of a False Negative Rate Due to Protein Polymorphisms – A polymorphism within a marker peptide could result in a false result. Consequently, it is important to determine the frequency of polymorphisms within marker peptides.

DETERMINING VARIANTS - Variant Tables for each body fluid marker protein were downloaded from gnomAD v2 and filtered for variants occurring within the target peptides for each protein. The Genome Aggregation Database (gnomAD) v2 is composed of 125,748 exomes and 15,708 genomes from diverse populations around the world[4] and provides rich resource for an investigation of the variation present in protein markers for blood, semen, and saliva.

RESULTS: While protein-coding regions, to varying degrees, exhibit constraint against deleterious mutation, variants that result in protein sequence difference do occur in the target regions of NYC OCME protein markers, largely at low population frequencies (Table 9). The only marker peptides with any variants with allele frequencies greater than 0.01 are LSEPAELTDAVK (KLK3 protein) and GQLPSGSSQFPHGQK (SEMG2 protein), both semen markers (Table 10).

Table 9: Summary of variants for body fluid marker peptides present in the gnomAD v2 database, with the number of variants and the maximum allele frequency of all variants listed for each peptide.

Protein	Marker Peptide	Number of SAAPs	Max. Allele Frequency
AMY1	ALVFVDNHDNQR	7	0.0023197
B3AT	ADFLEQPVLGFVR	11	0.0015242
B3AT	ASTPGAAAQIQEVK	1	0.0000159
B3AT	IPPDSEATLVLVGR	14	0.0000212
CYTT	ALHFVISEYNK	13	0.0003217
CYTT	ATEDEYYR	6	0.0000884
CYTT	SQPNLDTCAFHEQPELQK	13	0.0002863
HBA	FLASVSTVLTSK	4	0.0000322
HBA	TYFPHFDLSHGSAQVK	5	0.0000945
HBA	VGAHAGEYGAEALER	5	0.0000562
HBB	LLVVYPWTQR	6	0.0003077
HBB	SAVTALWGK	13	0.0002508
HBB	VNVDEVGGEALGR	14	0.0003573
HIS1	EFPFYGDYGSNYLYDN	17	0.0009689
KLK3	IVGGWECEK	7	0.0074349
KLK3	LSEPAELTDAVK	14	0.0825171
KLK3	SVILLGR	8	0.0001885
LEG1H	ESPGQLSDYR	6	0.0000080
SEMG1	EQDLLSHEQK	7	0.0000199
SEMG1	HLGGSQQLLHNK	9	0.0076851
SEMG1	SQIQAPNPK	3	0.0000080
SEMG2	GQLPSGSSQFPHGQK	14	0.1477283
SEMG2	GSISIQTEEK	9	0.0000080
SEMG2	LWVHGLSK	7	0.0000318

Note that while our assay includes three marker peptides for AMY1, two peptides had no recorded variants, and so do appear in the above table.

Table 10: Variant details for marker peptide variants with allele frequency greater than 0.01.

Protein	Peptide	Peptide Start	Peptide End	SAAP	VEP* Annotation	Allele Frequency	Homozygote Frequency
KLK3	LSEPAELTDAVK	126	137	p.Leu132Ile	missense_variant	0.0825171	0.0078413
SEMG2	GQLPSGSSQFPHGQK	30	44	p.Gln43Lys	missense_variant	0.1477283	0.0253004

1.3.4 Establishing Assay Quality Metrics Needed for Routine Monitoring - A system for regular evaluation and longitudinal monitoring of system performance and data quality control has been

established for NYC OCME's Molecular Serology assay. The two main components of the data quality management system are system suitability and quality control, or batch acceptance.

System suitability monitoring established for NYC OCME Molecular Serology uses a commercially available Peptide Calibration Mixture (PCM) of synthetic peptides that is run through the LC-MS system before and after every analytical batch. Measurements of peak area, retention time and full width at half maximum (FWHM) of the peptide peak width are recorded and evaluated. A set of benchmark experiments were used to establish target criteria for each of the measured PCM peptide metrics. Data collected from regular PCM runs is monitored both to ascertain whether the instrument system is performing to expectation prior to running an analytical batch, and to understand gradual shifts in performance of the instrument system over time.

Batch Quality Control, or batch acceptance criteria, consists of a system of standards and controls that include positive and negative control samples included in every analytical batch, a blank injection through the LC-MS system before every sample, and an internal standard (either Cytochrome C or heavy labeled synthetic marker peptides) spiked into every sample. A set of benchmark experiments was used to establish the acceptable performance for the internal standard, and for analytical signal above the blank. The positive and negative controls are evaluated using the same criteria for body fluid marker peptide identification applied to unknown samples. All standards and controls must meet established acceptance criteria in order for the batch results to be accepted. In addition, data from blanks, internal standards, and positive and negative controls are monitored and evaluated over time to understand gradual shifts or trends in performance of the assay.

1.4 Expected Applicability of the Research – The need to know the source of DNA in a forensic sample is becoming increasingly important as the amounts of DNA able to identify an individual can come from a relatively small number of cells. Consequently, a suspect may claim that their

DNA profile came from shed skin cells, primary or even secondary DNA transfer. Demonstrating, through a confirmatory test, that the source of a DNA STR profile is from blood, saliva or semen can provide important contextual evidence in investigations or at trials, resulting in fairer outcomes in the judicial system. We have developed, validated and brought into routine testing a confirmatory body fluid test at the NYC OCME. This test has been in use in our laboratory since July 2023. We have received inquiries about the test from other laboratories, and anticipate other laboratories will implement our test in the coming years.

References:

1. Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22, 209–212. doi: 10.2307/2276774.

2. Newcombe R.G. (1998). Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. Statistics in Medicine, 17, 857–872. doi:10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E.

3. Gessulat, S., Schmidt, T., Zolg, D.P. et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat Methods **16**, 509–518 (2019). https://doi.org/10.1038/s41592-019-0426-7

4. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. and Gauthier, L.D., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature, 581(7809), pp.434-443.

2) **Participants and Other Collaborating Organizations –** This work was carried out entirely at the New York City Office of Chief Medical Examiner.

3) Changes in Approach from Original Design and Reason for Change - Not applicable.

4) Outcomes – As defined in the NIJ's guidelines outcomes include: 1) Activities / Accomplishments, 2) Results and Findings, and 3) Limitations.

4.1 ACTIVITIES/ACCOMPLISHMENTS – This grant's overarching goal was to validate our confirmatory, proteomic body fluid assay for introduction into actual, routine, forensic casework. To achieve this end, we not only completed the work described in the grant but ensured that the assay was accredited by the national forensic accrediting agency ANAB (June 1, 2023) and approved by the New York State Commission on Forensic Science (June 9, 2023). The assay was introduced into casework at the NYC OCME in July 2023. As may be seen in 5.1 below, we have disseminated our results widely. Two members of my group, myself (Vice Chair) and Erin Butler (member), have been appointed affiliate members of the Forensic Proteomics Task Group started by the Organization of Scientific Area Committees (OSAC) for Forensic Science's Biology Scientific Area Committee (SAC). This group was constituted to set guidelines and standard for laboratories interested in employing proteomic mass spectrometry. One of the main focuses of this group is molecular serology.

4.2 RESULTS AND FINDINGS – The results of this grant have demonstrated the accuracy and reliability of proteomic mass spectrometry for confirmatory body fluid identification. Our results demonstrate that the assay is body fluid specific as, so far, there were no false positives in the hundreds of samples that we tested, and non-targeted body fluid samples (e.g. menstrual blood and vaginal fluid) were not misidentified. The assay also demonstrated the ability to distinguish primates from other species and was able to identify mixture of body fluids.

4.3 LIMITATIONS – Assay improvements can be made (e.g. increasing sensitivity) and these will be an ongoing process. One of our goals is to add additional body fluids to the assay, e.g. urine and menstrual blood. It is important to point out that this assay could not have been developed,

validated and put into production without grant support from NIJ. This work, based on that support will aid victims, help exonerate the wrongfully accused and make a fairer judicial system.

5) Artifacts – As per NIJ Guidelines, these include: 1) List of products (e.g., publications, conference papers), 2) Data sets generated, and 3) Dissemination activities

5.1 LIST OF PRODUCTS –

- PUBLICATIONS

1. Yang, H., Monier, S. A., Goldstein, M., Kaylee, Almubarak, I, Perez, T. Butler, E., and Siegel D., Identification of Marker Proteins of Menstrual Blood by Proteomic Mass Spectrometry, submitted: Int. J. Legal Med. April 2025.
2. The NYC OCME Molecular Serology Proteomic Body Fluid Validation (44 pages) is published on the NYC OCME public facing website: https://www.nyc.gov/assets/ocme/downloads/pdf/molecular_serology_validation.pdf
3. Manuals for the protocols used in Molecular Serology body fluid assay are published on the NYC OCME public facing website: https://www.nyc.gov/site/ocme/services/fbio-2025-molecular-serology.page, these include:
    a. Guidelines for Molecular Serology Body Fluid Assay
    b. Extraction
    c. Quantitation
    d. Digestion
    e. Liquid Chromatography & Mass Spectrometer Processing
    f. Liquid Chromatography & Mass Spectrometer Processing using Excel
    g. Analysis
    h. Molecular Serology Data Interpretation

- CONFERENCES PRESENTATION – Our work was presented at 17 conferences.

1. “Introduction of Proteomic Mass Spectrometry into Biological Forensics”, University of Rhode Island, Department of Chemical Engineering, October 27, 2022, Donald Siegel, invited oral presentation.

2. “Validation of a Confirmatory Forensic Body Fluid Assay Using Proteomic Mass Spectrometry” Pittcon NIJ Symposium: Innovations in Technologies to Advance Forensic Science Philadelphia, PA, March 19, 2023, Donald Siegel, invited speaker.

3. “Peptide Sequence Confirmation on MIDAS Data by Spectra Searching”, American Society of Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, Houston, TX, June 4 - 8, 2023, Heyi Yang, poster presentation.

4. “Validation of a Confirmatory Proteomic Mass Spectrometry Assay for Identification of Blood, Semen, and Saliva”, National Institute of Justice Forensic Science Research and Development Symposium: American Academy of Forensic Sciences, Orlando, FL, February 13-18, 2023, Erin Butler, oral presentation.

5. “Molecular Serology Forensic Body Fluid Identification using Proteomic Mass Spectrometry” Customer Presentation at the NYC OCME, June 27, 2023, Donald Siegel, oral presentation.

6. “Harnessing Proteomic Mass Spectrometry for Body Fluid Identification in Forensic Casework”, American Society of Crime Laboratory Directors Forensic Research Committee (ASCLD FRC) – Lightening Round, November 16, 2023, Donald Siegel, invited oral presentation (virtual).

7. “Validation of a Confirmatory Forensic Body Fluid Assay Using Proteomic Mass Spectrometry”, Forensic Technology Center of Excellence (FTCOE) R&D Webinar, December 07, 2023, Donald Siegel, invited oral presentation.

8. “Comparative Assessment Of Emerging Technologies For Body Fluid Identification” the American Academy of Forensic Sciences (AAFS), Colorado Convention Center, Denver, Colorado, Feb 19 - 24, 2024, Mirna Ghemrawi, The Center for Forensic Science Research & Education, presenter. (Our work was presented here by Dr. Ghemrawi in a comparison with other methods for body fluid identification.)

9. “Proteomic Mass Spectrometry - A Confirmatory Forensic Body Fluid Assay”, Mid-Atlantic Association of Forensic Scientists, Workshops: Current Practices and Developing Methods in Forensic Serology, Pittsburgh, PA May 8, 2024, Donald Siegel, invited speaker.

10. "A Confirmatory Forensic Body Fluid Assay Using Proteomic Mass Spectrometry", The Green Mountain DNA Conference, South Burlington, VT, July 22-24, 2024, Iyman Almubarak, invited speaker.

11. "Molecular Serology at OCME", Oral presentation for defense and prosecutors given at the New York City Office of Chief Medical Examiner, NY, NY, September 6, 2024, by Erin Butler.

12. "Perspective of a Research Scientist, Practitioner, and Policymaker", NIJ 2024 Research Conference, Advancing Justice Through Science, Pittsburgh, PA. September 16-18, 2024, Donald Siegel, invited speaker.

13. "A Likelihood Ratio Framework For Evaluating The Statistical Strength Of Body Fluid Identification Using Protein Markers", International Symposium on Human Identification, September 23-26, 2024, Erin Butler, oral presentation.

14. "A Confirmatory Protein-Based Mass Spectrometry Assay For Body Fluid Identification", Northeastern Association of Forensic Scientists, Atlantic City, NJ , October 18, 2024, Iyman Almubarak, invited speaker.

15. **"**Moving from Innovative R&D to Assay Production in a Regulated Environment/Establishing Quality Controls, System Suitability, & a Statistical Measure of Assay Confidence using Likelihood Ratio Framework", Pittcon NIJ meeting - Innovations in Technology to Advance Forensic Science, Boston, MA, March 3, 2025, Erin Butler, invited speaker.

16. **"**Bio-Assay Development and Validation - A Perspective from the New York City Office of Chief Medical Examiner" The Banbury Center Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory, NY, Microbial Forensics, May 12-15, 2025, Donald Siegel, invited speaker.

5.2 DATA SETS GENERATED – The primary data generated by this grant are mass spectrometry data collected from hundreds of body fluid samples. These samples include blood, saliva, semen, menstrual blood and vaginal fluid. Both neat (pure fluids) and mock forensic samples (body fluids applied to fabric, cotton swabs, etc.) were evaluated individually and in mixtures. Serial dilutions of neat and mock samples (to determine limits of detection) were also evaluated. **Data Archiving -** The following data types were generated during this project: Mass spectrometry-based proteomic profiling data (both qualitative and (semi)quantitative) in *.raw or *.wiff file formats. Standards developed by HUPO PSI will be used for dissemination of all data and results: mzML

(http://www.psidev.info/mzml_1_0_0%20) for mass spectra and spectral libraries, mzIdentML (http://www.psidev.info/mzidentml) for identified peptides. Mass spectrometry data were backed up and saved permanently by the NYC OCME and will be made available, upon publication, to the public using one or more of the following repositories: PRIDE, PeptideAtlas, Tranche, ProteomeExchange, MassIVE, and GPMdb. In addition, the NYC OCME Department of Forensics makes all validations, assays manuals including SOPs available on their public facing webpage https://www1.nyc.gov/site/ocme/services/fbio-protocols-for-str-analysis-manuals.page.

5.3 DISSEMINATION ACTIVITIES – As may be seen above in 5.1 LIST OF PRODUCTS, we have presented our data 16 times at scientific and forensic conferences and workshops, as well as lectures to New York City District Attorneys and the defense community in order to inform them of the assay implementation into routine casework and how it works. We have published one journal article and are currently in the process of completing two manuscripts on the developmental validation as well as final validation for submission to peer reviewed journals.

6) SUPPLEMENTARY DATA FOLLOWS

Supplemental Table 1 - BCA Protein Assay Quantitation of Body Fluids

Fold Dilution of Neat Body Fluid & Frequency of Detection

Semen

Dilution Factor	Percent Detectable (%)
1	100
2	100
4	89.8
8	55.9
10	89.8
20	61.9
32	0
64	0

Semen samples diluted at 32x and 64x are too dilute to read within the limits of the assay.

Saliva

Dilution Factor	Percent Detectable (%)
1	89.70%
2	38.70%
3	16%
4	0.94%
5	0%
10	0%
20	0.78%

Protein concentration in saliva is more variable than semen and blood, however undiluted saliva extraction shows the best sensitivity for the assay.

Blood

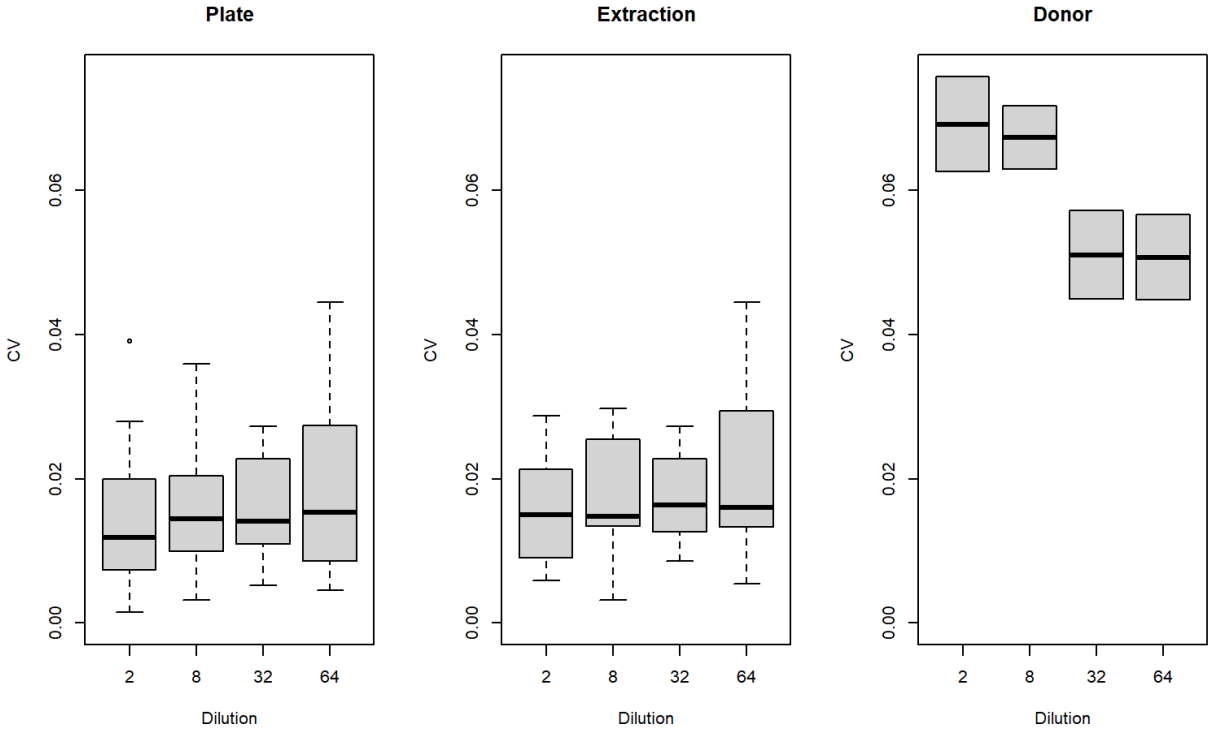
Dilution Factor	Percent Detectable (%)
5	100
10	100
20	100
40	100
200	8.33

Blood can be diluted at least forty times and still be detected.

Supplementary Tables 2-4 BCA Protein Quantitation Assay - Precision and Repeatability for Blood, Saliva and Semen

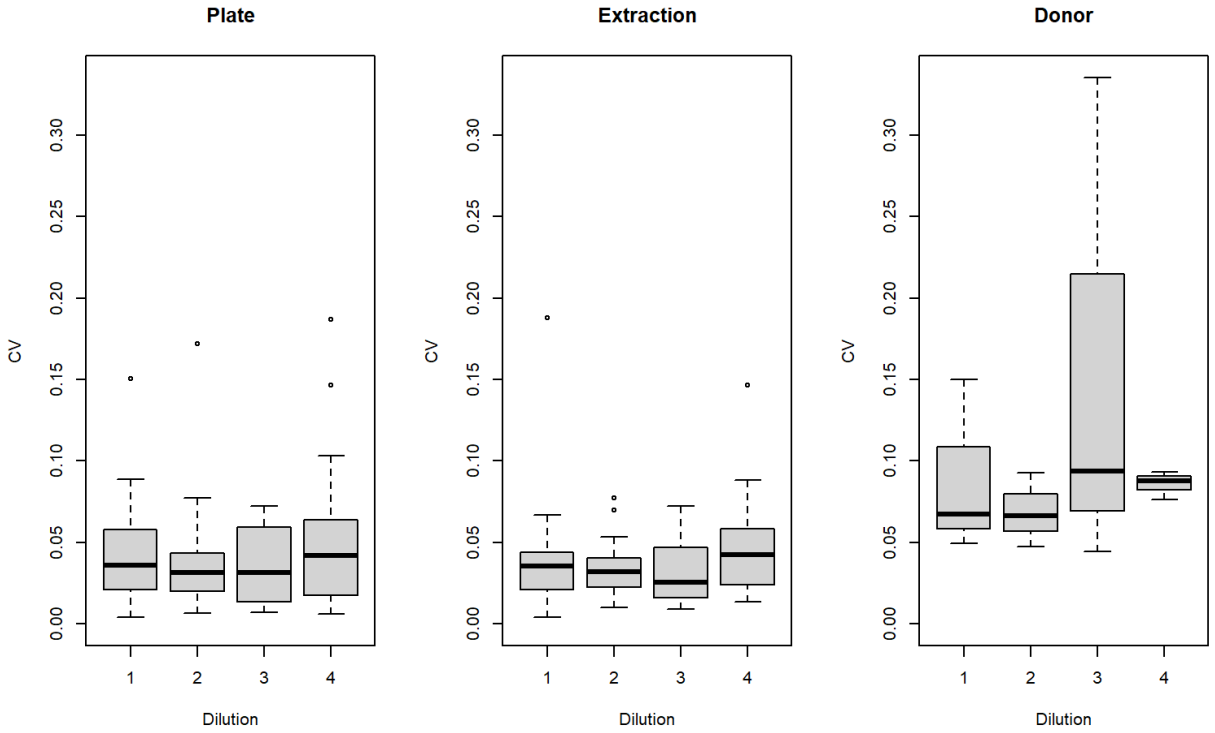
For samples where the same extraction and dilution were measured on the same plate at least 3 times, CVs were calculated for plate, extraction, and donor.

Supplementary Table 2 Semen



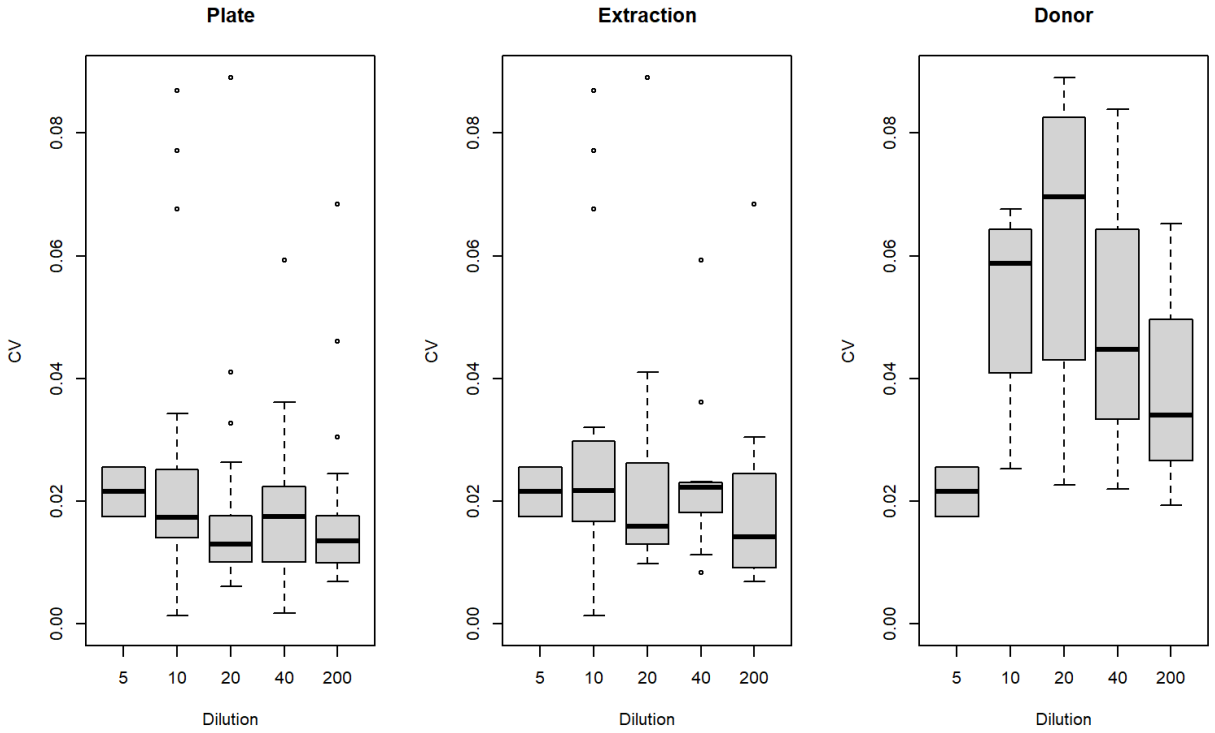
While absorbance CV increases for different extractions from the same semen donor, all are below 10%.

Supplementary Table 3 Saliva



Absorbance CV for saliva samples is low for repeat measurements of the same extracion, but variation increases more dramatically for different extractions from the same donor.

Supplementary Table 4 Blood



Similar to semen, blood absorbance CV, while higher for different extractions, is consistently under 10%.