Proteomics profiling of autologous blood and semen exosomes from HIV-infected and uninfected individuals reveals compositional and functional variabilities

Hussein Kaddour¹, Yuan Lyu¹, Jennifer L. Welch², Victor Paromov³, Sammed N. Mandape⁴, Shruti S. Sakhare⁵, Jui Pandhare⁶, Jack T. Stapleton², Siddharth Pratap⁷, Chandravanu Dash⁶, Chioma M. Okeoma¹*

¹Department of Pharmacology, Stony Brook University School of Medicine, Stony Brook, NY 11794-8651, USA
²Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA 52242-1109, USA
³Dept. of Microbiology, Immunology and Physiology, Meharry Medical College, Nashville, TN 37208
⁴College of Medicine, The University of Arizona Health Sciences, Tucson, AZ 85721
⁵University of North Texas Health Science Center, Fort Worth, TX 76107
⁶Department of Biochemistry and Cancer Biology, Meharry Medical College Nashville, TN 37208
⁷School of Graduate Studies and Research, Meharry Medical College Nashville, TN 37208

*Corresponding author, Chioma M. Okeoma, Email: chioma.okeoma@stonybrook.edu

Running title: Analysis of blood and semen exosome proteomes

Key words: Exosomes, HIV, blood, semen, proteins
Abbreviations: SMPP, Systmatic Multifactorial Proteomic Profiling; BE, Blood exosomes; SE, Semen exosomes; EFBP, Exo-free blood plasma; EFSP, Exo-free seminal plasma; MVEs, Multi-vesicular endosomes; ExProteins, Extracellular proteins; ExNucleic acids, Extracellular nucleic acids; DPP, Differentially present proteins; SpC, Spectral counting; AUC, Area under the curve; MudPIT, Multidimensional protein identification technology;

Abstract

Blood and semen are important body-fluids that carry exosomes for bioinformation transmission. Therefore, characterization of their proteomes is necessary for understanding body-fluid-specific physiologic and pathophysiologic functions. Using systematic multifactorial proteomic profiling, we characterized the proteomes of exosomes and exosome-free fractions from autologous blood and semen from three HIV-uninfected and three HIV-infected participants (total of 24 samples). We identified exosome-based protein signatures specific to blood and semen along with HIV-induced tissue-dependent proteomic perturbations. We validated our findings with samples from 16 additional donors and showed that unlike blood exosomes (BE), semen exosomes (SE) are enriched in clusterin. SE but not BE promote Protein•Nucleic acid binding and increase cell adhesion irrespective of HIV infection. This is the first comparative study of the proteome of autologous BE and SE. The proteins identified may be developed as biomarkers applicable to different fields of medicine, including reproduction and infectious diseases.

Introduction

Intercellular communication in complex biological systems, such as blood and semen occur through “horizontal” transfer of bioactive RNA and protein associated with extracellular vesicles, such as exosomes. It is generally agreed that cells from various multicellular organisms, including humans secrete bioactive exosomes into the extracellular milieu [3-7]. Exosomes are membrane
nanovesicles that originate from the inward budding of endosomal membranes into intracellular multi-vesicular endosomes (MVEs). Exosomes within the MVEs are released into the extracellular milieu following fusion of MVEs with the plasma membrane. As a result, various body fluids contain tissue-specific extracellular molecules (exNucleic acids, exProteins) that play site-specific biological roles [4]. A myriad of exNucleic acids and exProteins are encased within exosomes [5, 6, 8-11] as exosomal cargos. Through their cargos, exosomes from cell culture and body-fluids modulate host responses in a variety of physiological and disease conditions, including cancer and various infections [4, 12].

Interest in exosomes as vehicles of bioinformation transfer has motivated studies of exosome cargo bio-distribution. Exosome-specific cargo may be used as signatures for various physiologic and pathophysiologic conditions. The body-fluids - blood and semen are important vehicles for transmission of various viral agents, such as ZIKA virus, herpes simplex virus 1 and 2 (HSV-1 and -2), cytomegalovirus (CMV), and human immunodeficiency virus (HIV-1 and -2, [HIV]). HIV for example relies heavily on the host cellular machinery to replicate in a tissue-specific manner. Although viral RNA exists in both blood and semen [13-18], local environmental factors may reprogram the transcriptome and proteome of blood and semen in a body-fluid specific manner. Since the cargo, composition, and function of exosomes mirror the status of the producer cell, it stands to reason that infection with HIV may influence exosome composition in a body-fluid specific manner. Indeed, exosomes from blood and semen have intriguingly shown variability in function [5, 11, 19-21]. Exosomes isolated from semen but not blood of HIV-negative human donors contain protective factors that endow an anti-HIV phenotype [5, 11, 19-21].

Significant portions of body-fluids are made up of soluble and vesicle-bound proteins [22, 23]. However, the repertoire of extracellular proteins (exProteins, in particular, exosomal proteins) in disease-relevant tissues has been neither systematically identified nor qualitatively characterized. One major confounding factor of exosome research is that exosomal cargos are heterogeneous and mirror the landscape of the producer cells. Although different mass
spectrometry-based approaches have been used to profile exosomal proteins [24-35], none has combined multiple techniques into one approach to profile exosomal proteins both within and between donors from different clinical groups. Our major objective here is to profile the repertoire of exosomal proteomes in different physiological/pathophysiological states and body-fluid types, and to uncover the consequences of HIV infection in exosome composition. While difficult to come by, it is important to use autologous body-fluids in such analysis to promote rigor and reduce confounding errors introduced by variation from different donors.

To this end, we combined discovery proteomics and multifactorial analysis of autologous blood and semen to profile exProtein signatures of three HIV-negative (HIV-) and three HIV-positive virus-suppressed (HIV+) participants, and validated our findings with 16 unpaired blood and semen samples. Considerable inter-individual variability in protein profile and compartmentalization between the body-fluids was observed, underscoring the importance of individual autologous donor studies. However despite the variations, differentially present proteins (DPP) in blood and semen were identified and distinct pathway signatures were revealed. Furthermore, the functional implications of the different protein profiles were demonstrated by analyzing patterns of Protein•Nucleic acid binding and cell adhesion—Protein•Protein interactions of BE and SE.

Our study provides the first exosomal protein compartmentalization and function in autologous blood and semen from two different clinical groups. The proteins identified have potential biomarker value which may be developed for clinical decision-making processes applicable to different fields of medicine, such as reproduction and infectious diseases.
Experimental procedures:

Ethics: All experiments in this study were completed according to University regulations approved by The University of Iowa and Stony Brook University Institutional Review Boards (IRB). All participants were adults. They provided written informed consent for both blood and semen samples, and all laboratory personnel were blinded to clinical data. Clinical characteristics were obtained through review of medical records (Supplemental Table 1). The University of Iowa Hospitals and Clinics clinical laboratory measured plasma viral load (RNA) using the Roche Cobas method, and CD4+ T-cell counts using flow cytometry as previously described [36]. Standard operating procedures [4, 5, 11, 19, 21, 36, 37] was used for blood and semen collection.

Experimental Design and Statistical Rationale

The current study was designed with a focus on identifying the differences between blood and semen exosomes proteomes because semen but not blood exosomes inhibit HIV infection of target cells in vitro. To this end, six participants, three HIV-negative (HIV-) and three HIV-infected (HIV+) participants donated autologous blood and semen used in proteomics analysis. We analyzed exosomes from every donor sample individually as independent biological replicates to capture random biological variation. In parallel, we also analyzed the paired exosome-free plasma (n = 12) as independent biological replicates used as control to identify exosome specific proteins from blood and semen. Additional 16 participants (six HIV- and 10 HIV+) donated blood and semen, whose exosomes and exosome-free plasmas were pooled and used for data validation (Supplemental Table 1).

Isolation of exosomes: Exosomes were purified from blood plasma or semen, as previously described [37, 38]. Briefly, plasmas were differentially centrifuged at 500 x g for 5 min, 2000 x g for 10 min and 10,000 x g for 30 min, to remove cells, cellular debris, and large microvesicles, respectively. Clarified supernatants were then precipitated using ExoQuick reagent (SBI catalog
#EXOQ5A-1) at 4°C overnight. Exosomes were pelleted by centrifugation at 1,500 × g for 30 min, and supernatant was collected as exosome-free plasma. Exosome pellets were centrifuged again at 1,500 × g for 5 min, the leftover supernatant was discarded and the pellets were resuspended in PBS (1/10th of volume). Exosome and exosome-free plasma proteins were quantified by the Bradford method (BioRad), aliquoted and stored at -80 °C until use.

**ZetaView NanoTracking Analysis (NTA) of exosomes:** The concentration and size distribution of exosomes were determined by NTA using ZetaView (PMX110, Particle Metrix). The system was calibrated using 102 nm standard silica beads. Exosome samples were diluted between 20,000 and 100,000 times and measurements were acquired using ZetaView software v8.04.02. Particle concentrations were reported per mL of plasma so that a comparison between samples from different clinical groups was made possible. Measurements of 11 positions were performed in triplicate and outlier positions were automatically excluded.

**Exosome Sample Preparation:** Prior to the LC-MS analysis, protein samples were denatured in 8M urea and 50 mM Tris-HCl, pH 8.0, reduced with 10 mM TCEP for 60 min at RT, alkylated with 2 mM iodoacetamide for 60 min at RT, and then diluted to 2M urea with 50 mM Tris-HCl, pH 8.0. Two micrograms of Trypsin Gold (Promega) was added for overnight digestion (18 h, 37° C), and then the tryptic peptides were immediately desalted using Pierce C18 spin columns (Thermo Fischer Scientific) at RT. Peptides were eluted with 80% acetonitrile and 0.1% formic acid (FA), and dried completely on a SpeedVac Concentrator.

**Database Search Parameters And Acceptance Criteria For Identifications:**

**Peak Lists and Search Engine Parameters:** Peak lists, protein identifications and database searches were conducted using BSI PEAKS Studio software employing search engine version 8.5 (Bioinformatics Solutions Inc., Waterloo, Ontario Canada). For label-free quantitation (LFQ),
we employed the Q module of BSI PEAKS software which uses expectation - maximization algorithms on the eXtracted Ion Chromatograms (XIC) of the three most abundant unique peptides of a protein to calculate the Area Under the Curve (AUC) [39].

**Sequence Databases:** We constructed an in-house sequence database using the Swiss-Prot UniProt Human non-redundant, manually annotated database (up000005640, version release 2018_04) as the reference database (https://www.uniprot.org/uniprot) [40]. This consisted of 20,303 annotated human proteins, to which the ten HIV-1 proteins (RefSeq accession NC_001802.1, version release 13-Aug- 2018) and nine HIV-2 (RefSeq accession NC_001722.1, version release 13-Aug- 2018) viral protein sequences were concatenated in order to construct our final sequence database. Enzyme specificity was fully tryptic with maximum 3 missed cleavages and maximum 1 non-specific cleavages. Modifications used were Carbamidomethylation (57.02) as fixed and Oxidized Methionine (15.99) as variable. Parent mass error tolerance was set as 20.0 ppm, and fragment mass tolerance set to 0.5 Da.

Known contaminants to be excluded were identified and removed using the common Repository of Adventitious Proteins (cRAP) database version 1.0, release 2012.01.01 (https://www.thegpm.org/crap/). This is a listing of common laboratory proteins, including bovine serum albumin (BSA) and trypsin precursors, non-sample lab contaminants from dust and human sample handling and molecular weight standard proteins.

**Threshold score/expectation value:** The BSI PEAKS peptide score (-10lgP) was used for significance score of detection was for all peptide-spectrum search results. This is a derived score from the peptide-spectrum match (PSM) p-value. The protein level PEAKS score is the weighted sum of the -10lgP PEAKS peptide scores. A PEAKS protein score of >= 20 was used as the significance threshold for all database search results. For the label-free quantitation (LFQ), an additional threshold of XIC AUC of the 3 most abundant peptides of a protein to be > 1e5 [41].
**False Discovery Rates at Peptide and Protein levels:** FDR for the low-stringency protocol was set to 0.1% at the peptide-spectrum match (PSM) level and corresponded to FDR of 2% at the Peptide Sequence level. For the high-stringency protocol, a 1% Peptide Sequence level threshold was used, resulting from a PSM FDR < 0.1%

**Nano-LC-MS/MS Analysis:** Peptides were resuspended in 5 μL of 0.5% FA and loaded onto a 3-phase MudPIT column (150 μm x 2 cm C18 resin, 150 μm x 4 cm strong cation exchange SCX resin, filter union, and 100 μm x 12 cm C18 resin) as described previously [42]. A 10-step MudPIT (0 mM, 25 mM, 50mM, 100 mM, 150 mM, 200 mM, 300 mM, 500 mM, 750 mM, and 1000 mM ammonium acetate, each salt pulse followed by a 120-min acetonitrile gradient 5-50% B [Buffer A: 0.1% FA; Buffer B: 0.1% FA in acetonitrile]) was executed for LC-MS anal ysis using an Eksigent™ AS-1 autosampler and Eksigent™ nano-LC Ultra 2D pump online with an Orbitrap LTQ XL linear ion trap mass spectrometer (Thermo Finnigan) with a nanospray source. MS data acquisition was done in a data-dependent 6-event method (a survey FTMS scan [res. 30,000] followed by five data-dependent IT scans for five consequent most abundant ions). The general mass spectrometric settings: spray voltage, 2.4 kV; no sheath and no auxiliary gas flow; ion transfer tube temperature, 200 °C; CID fragmentation (for MS/MS), 35% normalized collision energy; activation q = 0.25; activation time, 30 ms. The minimal threshold for the dependent scans was set to 1000 counts, and a dynamic exclusion list was used with the following settings: repeat count of 1, repeat duration of 30 s, exclusion list size of 500, exclusion duration of 90 s.

**Data mining and visualization:** KEGG pathways and GO terms were determined using WEB-based Gene SeT AnaLysis Toolkit 2019 [43]. CLU PPI network was visualized using STRING database V11 [2]. ANXA4 and CCT6A PPI analysis was performed using Molecular Interaction Search Tool, MIST [44], https://fgrtools.hms.harvard.edu/MIST/, February 11, 2019). Heatmaps were visualized using GraphPad Prism (v8.01). Clustering Heatmaps and square pairwise correlation matrices were drawn using heatmapper [45]. The clustering method used was the
average linkage with Euclidean distance measurement applied to both rows and columns. Venn diagrams were obtained using Venny platform (v2.1) [46].

**Acetylcholinesterase assay (AChE):** Exosome-specific AChE enzymatic activity was measured as previously described [20]. Briefly, 50 μg BE or SE from HIV- (n=3) and HIV+ (n=3) donors were lysed in 2% Triton X-100 and mixed with a 1:1 volumetric ratio of 1.25 mM acetylthiocholine chloride (Sigma-Aldrich) and 0.1 mM 5,5’-Dithiobis 2-nitrobenzoic acid (Sigma-Aldrich). Absorbance was read at 450 nm for 30 min at 37°C at 5 min intervals. Equivalent volume of PBS was used as AChE negative control. Data are reported as the individual mean from triplicate measurements. Error bars are SD.

**Cells and virus:** TZM-bl (NIH AIDS Reagent Program) and 293T (ATCC) cells were maintained in DMEM (Gibco-BRL/Life Technologies) containing 5% exosome-depleted FBS (Gibco), 100 U/ml penicillin, 100 μg/ml streptomycin, 100 μg/ml sodium pyruvate, and 0.3 mg/ml L-glutamine (Gibco-BRL/Life Technologies). SUPT1 and Jurkat cells (NIH AIDS Reagent Program) were maintained in RPMI (Gibco-BRL/Life Technologies) containing 5% exosome-depleted FBS (Gibco), 100 U/ml penicillin, 100 μg/ml streptomycin, 100 μg/ml sodium pyruvate, and 0.3 mg/ml L-glutamine (Gibco-BRL/Life Technologies). HIV-1 NL4.3 virus was produced from NL4.3 plasmid (NIH AIDS Reagent Program). 293T cells were transfected with pNL4.3 by Lipofectamine 2000 (Invitrogen) in antibiotic-free media as previously described [5, 19, 20]. Virus titers were determined by TZM-bl relative light units (RLU).

**HIV LTR promoter activation:** 200 μg BE or SE from HIV- (n=3) and HIV+ (n=3) donors were added to TZM-bl cells (NIH AIDS Reagent Program) in triplicate for 24 hours. TZM-bl cells contain HIV-1 LTR that expresses luciferase when activated. PBS was added to cells as vehicle negative control, and HIV-1 NL4.3 virus was added as positive control. After treatment for 24 hours, treatment was removed and cells were lysed and measured for luciferase reporter activity.
(Steady-Glo, Promega). Data are reported as the group mean (HIV- or HIV+) from individuals measured in triplicate. Values are reported in relative light units (RLU). Error bars are SD.

**HIV RNA evaluation:** 200 μg BE or SE from HIV- (n=3) and HIV+ (n=3) donors were added to SUPT1 cells (NIH AIDS Reagent Program) in duplicate for 24 hours. PBS was added to cells as vehicle negative control, and HIV-1 NL4.3 virus was added as positive control. After treatment for 24 hours, treatment was removed and cells were washed in PBS. RNA was extracted from cells using RNeasy kit with DNase treatment per manufacturer’s instructions (Qiagen). Equivalent concentrations of RNA (1 ng/μl) were used for cDNA synthesis (ABI). PCR analysis was completed with 500 ng cDNA and gene specific primers to amplify GAPDH and HIV-1 Gag-Pol using SYBR green master mix (ABI/Life Technologies) and 7500 qRT-PCR (ABI). qRT-PCR was analyzed by relative quantification where transcripts were normalized to corresponding GAPDH value. Normalized values were then quantified by fold change to HIV-uninfected cells set at 1. Data is reported as the group mean fold change (HIV- or HIV+) from individuals measured in duplicate. Error bars are SD. GAPDH F: 5′-CCCCTTCATTGACCTCAACTACA-3′ [47]. GAPDH R: 5′-CGCTCCTGGAGGATGGTGAT-3′ [47]. HIV-1 Gag-Pol F: 5′-TTCTTCAGAGCAGACCAGACG-3′ [48]. HIV-1 Gag-Pol R: 5′-GCTGCCAAAGAGTGATCTG-3′ [48].

**Western blot:** 50 μg BE or SE from HIV- (n=3) or HIV+ (n=3) donors as well as their exosome-free fractions were separated on 4-20% SDS-PAGE and the proteins were transferred to a PVDF membrane which was blocked with 5% BSA in TBST. The membrane was incubated overnight with primary antibodies as follow: clusterin (VWR), CD9 and CD63 (Developmental Studies Hybridoma Bank (DSHB), CD81 and FN1 (Novus Biologicals), HSP70 (R&D Systems), GAPDH, β-ACTN, ANXA4, CCT6A and ALB (Proteintech Group). After TBST washes, the membranes were incubated with fluorescently labelled secondary IgG antibodies IRDye® 800 (LI-COR) imaged using the Odyssey infrared imaging system (LI-COR).
Electrophoretic Mobility Shift Assay (EMSA): EMSA was conducted as previously described [19]. Briefly, 50 µg BE or SE from HIV- (n=3) or HIV+ (n=3) donors, as well as with their corresponding exo-free fractions, were lysed with 0.1% Triton X100 in EMSA binding buffer (20 mM HEPES pH 7.9, 50 mM KCl, 0.1 mM EDTA, 5% (v/v) glycerol, 0.2 mg/mL BSA and 1 mM DTT). 1 µg poly dl-dC competitor (LightShift, ThermoFisher) was added and the mixture was incubated on ice for 30 minutes. 1.5 fmol of Cy5-labelled NF-kB p65 probe or scrambled primer (5’Cy5-CTCGTATGTTGTGTGGAATTGTGAGCGGAT) was added and the mixture was further incubated on ice for 30 minutes. 2 µL of 10X orange loading dye was finally added and the mixture was loaded on a 5% 20cmx22cm non-denaturing PAGE. Gel was imaged using the Odyssey infrared imaging system (LI-COR). Experiment was done in duplicate with similar results.

Cell clustering and adhesion: The wells of a 96-well tissue-culture plate were coated with 50 µg BE or SE from HIV- (n=3) or HIV+ (n=3) donors, as well as with their corresponding exo-free fractions, in 100 µl PBS overnight at 4 °C. 100 µl PBS were used as vehicle negative control. The next day the exosomes were removed and the wells were blocked with sterile 2% BSA solution for 2 hours at room temperature. Subsequently, BSA was removed, wells were washed once with PBS, and 0.05 M cells of U937 monocytic cells or Jurkat T cells were added to each well and cultured overnight at 37 °C, 5% CO2. The following day, cells were imaged for visual analysis of cell clustering. Then, non-adherent cells were removed by PBS washes (3 times). Adherent cells were imaged using an automated microscope capable of automatic cell counting (LionHeart FX, Biotek). Error bars represent SD of 5 replicates. Experiment was repeated three times with similar results.

Statistical analysis: Differentially present proteins were determined using the 2-way ANOVA test by controlling the FDR (original FDR method of Benjamini and Hochberg). Differences were considered to be significant when Q < 0.05. The linear regression between the SpC data and AUC data was performed on the average of the values for each identified protein. The Bias and
agreement of the Bland-Altman plot were determined by showing the 95% limits of agreement. The adhesion events obtained using an automated microscope were compared with one-way ANOVA (Sidak’s test). The mean of each treatment was compared to the mean of PBS vehicle. HIV effects on SE and BE adhesive capabilities was assessed using two-tailed unpaired Student’s t test (with Welch’s correction). All calculations were performed using GraphPad Prism (v8.0.1).

Results

Pipeline for systematic multifactorial proteomic profiling (SMPP): We aimed to identify exProteins that are compartmentalized in different fractions of the body-fluids to identify differences between HIV- (physiological) and HIV+ (pathophysiological) donor blood and semen and to decipher what perturbations in protein cargo will be associated to HIV infection. For this purpose, we designed a pipeline specifically focusing on controlling errors and noise while at the same time avoiding a priori introduction of many constraints into the analysis. Our pipeline, which we call SMPP (systematic multifactorial proteomic profiling) is composed of four measures (Fig. 1a): i) donor variability (reproducibility); ii) tissue and compartment specificity (uniqueness); iii) protein present, measured by presence of at least two unique peptide peaks using spectral counting (presence, SpC); and iv) protein abundance significantly changed, measured by relative (pathophysiological-to-physiological) peptides’ parent ion intensities calculated via the parent extracted ion chromatogram (XIC) area under the curve from the mass spectra data (relative abundance, AUC). The idea behind combining these measures into one ensemble is based on the assumption that the measures will generate high confidence exProtein biosignature while minimizing errors. In particular, this process may minimize false positives in exProtein cataloging and type 2 errors that may be missed with less sensitive methods due to individual variability. We applied the SMPP protocol to separate clinically-guided autologous blood and seminal plasma
into (i) blood exosomes—BE, (ii) semen exosomes—SE, (iii) exosome-free blood plasma—EFBP, and (iv) exosome-free seminal plasma—EFSP (Fig. 1b). Exosomes size and concentration were characterized using Nano Tracking Analysis (NTA) (Fig. 1c), and by enzymatic activity of the exosome-associated marker, acetylcholinesterase (AChE) (Fig. 1d). As expected, NTA and AChE analyses revealed donor dependent variabilities in exosome concentration, size, and enzymatic activity.

**Elucidation of global exProtein landscape in blood and seminal plasma:** To elucidate the repertoire of exProteins in blood and seminal plasma irrespective of physiological and pathophysiological conditions of study participants, we used multidimensional protein identification technology—MudPIT [42] to analyze the 24 proteomes and to compare the relative abundance of proteins from BE, EFBP, SE, and EFSP fractions. A hierarchical approach (Fig. 1b) was used to map the individual proteome of the exosomes and their paired exosome-free counterparts for each donor in the two clinical groups. In an initial analysis, we used the spectral counting (SpC) method with two cutoffs at the peptide level: The first less-stringency SpC ($\leq 2\%$ FDR, protein detection significance threshold ($P$) of $-10\text{lg} P \geq 20$ and $\geq 1$ unique peptide) yielded 851 proteins and the second higher-stringency SpC ($\leq 1\%$ FDR, protein detection significance threshold of $-10\text{lg} P \geq 50$ and $\geq 2$ unique peptides) yielded 625 proteins. To define counting bias associated with proteins falsely discovered by the SpC method, we also used the area under the curve (AUC) method that accounts for peptide peak intensity (cutoff = $10^5$ a.u.) using PEAKS DB studio (version 8.5) [49]. This method resulted in an AUC list of 605 proteins. A 3-way Venn diagram of these three lists revealed data reproducibility between AUC and stringent SpC methods with 587 ($571 + 16$) protein overlap, accounting for 94% and 97% of the AUC and the stringent SpC lists, respectively, and an additional 233 proteins for the loose SpC list (Fig. 2a). Therefore, combination of these SpC and AUC may provide the analytical rigor suitable for our
complex data set. Next we compared our lists to the Top-100 most abundant exosomal markers from the ExoCarta exosome proteins [50], lipids and RNAs database (http://www.exocarta.org/) and found 69 proteins to be common, 54 of which were present in all lists, indicating that any of our filtration methods could be reliable (Fig. 2b, Supplemental Fig. 1). Furthermore, comparison of our dataset with similar published studies showed common proteins in both BE (Supplemental Fig. 2) and SE (Supplemental Fig. 3), increasing the confidence in our lists.

For discovery purposes, we used the 851-list coupled with Venn-fractionation to generate global representation of proteins present in BE, EFBP, SE, and EFSP for the two clinical groups (Fig. 2c-d). The results show a balanced protein compartmentalization in exosome fractions and their corresponding exosome-free plasma. However, SE fractions showed consistently higher number of proteins in HIV- and HIV+ specimens, suggesting that either semen may generally contain more distinct proteins compared to blood or contain less highly abundant proteins that mask other proteins for the LC-MS analysis (e.g. immunoglobulins).

Prioritization of proteins for discovery of protein targets for downstream analysis: Although the SpC-based quantification method is widely used for discovery, the method has limitations as a label-free semi-quantitative assessment of protein abundance [51]. Indeed, our analysis showed that sample clustering based on the SpC data segregated blood versus semen samples but did not efficiently discriminate between HIV- versus HIV+ samples (Fig. 3a). We found that the SpC method is biased toward high abundance proteins and inefficient in quantifying the low-abundance proteins (supplemental Fig. 4a,b). In contrast, the AUC method is assumed to be a better approach to label free quantitation of proteins identified by mass spectrometry [52, 53]. However, the AUC method did not reach satisfactory sample clustering in our dataset (Fig. 3b). The AUC method was biased toward the low-abundance proteins, where proteins with no spectral counts showed quantifiable (>10^5 a.u.) AUC (supplemental Fig. 4c). To identify a less error prone
quantitation method, we analyzed the correlation between SpC and AUC methods. Linear regression analyses show that SpC correlated to AUC, but to a limited extent ($R^2 = 0.56$) (Fig. 3c). The Bland-Altman plot [54] showed a proportional bias where the ratio between the two methods (SpC/AUC) increased with the mean of their values (Fig. 3d). This proportional bias indicates that the methods do not agree equally through the range of measurements, corroborating the inherent bias in both the SpC (Fig. 3a) and AUC (Fig. 3b) methods. Furthermore, square pairwise correlation matrices for the proteins in each method showed apparent visual similarities, with SpC being slightly more resolved (supplemental Fig. 4d,e). Altogether, these analyses indicate that, although derived from the same original MS data, SpC and AUC methods of analyses are complementary but not interchangeable, and analyses based on one or the other may result in unintended anomaly. For a more robust analytical process, we integrated the two methods by adjusting the AUC list to the SpC list, designated as SpC_Adjusted_AUC. Particularly, we nullified AUC values for proteins with undetectable spectral counts for each sample. This adjustment resulted in a new list of 587 proteins (proteins common to 625 stringent SpC and 605 AUC lists) where the samples readily clustered in HIV- vs HIV+ groups, revealing visual blood and semen proteome differences, as well as, HIV-induced and donor-dependent variabilities (Fig. 3e). For analyses in subsequent sections, we used the stringent SpC list to rule out the presence or the absence of a protein in a particular sample, and the SpC_Adjusted_AUC list for protein abundance.

Profiling exProteins common to human blood and semen irrespective of HIV infection and independent of compartment localization: For identification of proteomic profiles common to blood and semen irrespective of compartment localization (exosome and exosome-free) and donor physiologic or pathophysiological states (HIV- and HIV+), we used the step-wise Venn-filtration analysis of SMPP. Four-way Venn analysis depicts donor-dependent variabilities in protein compartmentalization. For each donor, we identified proteins that are common (bold fonts,
four-way Venn) to both exosomes and exosome-free compartments. A three-way Venn-filtration identified 22 and 8 exProteins that are commonly present in both the blood and semen of HIV- and HIV+ donors, respectively (Fig. 4a). Some identified proteins might be involved in adaptive immunity, including IGHG2, IGKC, IGHG4; and male fertility, such as semenogelin, lactoferrin, and clusterin. A two-way Venn-filtration between blood and seminal proteomes from HIV- and HIV+ participants revealed that all the eight proteins common to HIV+ are present in HIV- donors (Fig. 4b) at varying levels (Fig. 4c,d). This analysis therefore produced eight (ALB, SERPINA1, IGHG2, IGKC, IGHG4, CLU, IGHA1, IGHA2) proteins that are common to blood and semen irrespective of compartmentalization or HIV infection, although some levels of compartment-dependent enrichment was observed. Four proteins (ALB, CLU, IGHG2, and IGHA2) were enriched in a body-fluid-dependent manner. ALB and CLU levels were significantly higher in semen while IGHG2 and IGHA2 were significantly higher in blood (Fig. 4c,d). On the other hand, two (IGKC and IGHG4) of the eight proteins showed quantifiable AUC in all fractions without significant differences between the clinical groups (Fig. 4c,d).

One of the universal proteins is the immunoglobulin kappa chain (IGKC), which in cancer is used as a single, robust immune marker of cancer metastasis-free survival and response to chemotherapy [55]. Interestingly, the common immunoglobulins—IGKC and IGHG4, which probably originated from plasma cells and/or plasma blast are present not only in blood but also in semen. The fact that these proteins are not compartmentalized and not perturbed by donor HIV infection, indicate that they are conserved and could serve as universal internal marker for body fluid proteome. Contrary to the profile of IGKC, enrichment of clusterin (CLU), an extracellular chaperon glycoprotein, is compartment-dependent. In our dataset, CLU is highly enriched in HIV+ and HIV- SE (Fig. 4c). This observation is in line with the fact that the concentration of CLU in semen ranges from 0.4–15.0 mg/ml, which is about 20-fold higher than that of blood [56, 57]. CLU has previously been shown to be present in seminal secretions [58] with functions related to reproduction including damaging oxidative reactions [59], protein precipitation [60], agglutination
of abnormal spermatozoa [61], control of complement-induced sperm lysis [62], and infectious diseases, such as binding to DC-SIGN [63] and inhibition of HIV infection [64]. Since CLU is important in male fertility and HIV inhibition, and the protein is mostly enriched in SE from all six donors irrespective of their HIV status, CLU can be considered a potential SE signature. Indeed, CLU was previously described as a biomarker of fertility in stallions [65] and humans [66]. Protein-protein interaction (PPI) network analyses of CLU using STRING database (http://string-db.org/) [1, 2] identified highly-interconnected network of proteins with CLU (Fig. 4e). Proteins identified to have both known and predicted interaction with CLU include ALB, which is present and highly enriched in semen compared to blood (Fig. 4c,d). GO Terms associated with CLU ranged from negative regulation of protein metabolic process, protein binding, and growth factor binding, while pathways in cancer are amongst the top 5 KEGG pathways (Table 1). Protein-based validation experiments using 16 unpaired pooled BE and SE samples confirmed enrichment of CLU in the seminal fractions (Fig. 4f). These analyses identified exProteins that may serve as a tool for exosome research and highlight the need for further experimental assessment of the role of identified exosome-associated proteins in human physiology and pathophysiology.

Blood and semen exProtein compartmentalization in exosome-free plasma and exosomes of HIV-negative and HIV-positive individuals: To assess the extent of exProtein enrichment in blood and seminal plasma from different clinical groups, we performed a comparative analysis of proteins in different compartments (exosomes and exosome-free plasma) using the spectral counting method [67] with the stringent 625 SpC list. A two-way Venn diagram of proteins in each compartment for individual participants revealed donor-dependent variation in protein compartmentalization between exosomes and exosome-free plasma from HIV- (Fig. 5a) and HIV+ (Fig. 6a) donors. After filtering out exosome-free plasma proteins at the donor level, a three-way Venn-filtration of unique exosome proteins from all donors identified common exProteins in exosomes present in every donor (Fig. 5b, bold font) and (Fig. 6b, bold font). A similar analysis
was performed for the exosome-free fraction, but few proteins withstood the analysis (Supplemental Fig. 5). These data indicate that majority of the exProteins are compartmentalized in exosomes, and that this trend is independent of HIV infection or body-fluid type. Furthermore, we used the AUC method with the SpC-adjusted_AUC to obtain the quantitative intensity profile of differentially present proteins (DPP) per donor. Two-way ANOVA (P value < 0.05) with the original FDR method of Benjamini and Hochberg resulted in 320 DPP (102 BE-enriched and 218 SE-enriched) in HIV- (Fig. 5c,d, supplemental Table 2) and 337 DPP (91 BE-enriched and 246 SE-enriched) in HIV+ (Fig. 6c,d, supplemental Table 3). KEGG Pathway enrichment analysis of the DPP and pathways with significance values of P < 0.05 were visualized. Top 10 pathways in HIV- (Tables 2-3) and HIV+ (Tables 4-5) exosome proteome show that BE is enriched for Terms relating to complement activation cascade, response to infective agents, as well as, vitamin and fat digestion and absorption. On the other hand, SE is enriched for Terms relating to lysosome, other glycan degradation, biosynthesis of amino acids, metabolism, and protein processing. A 4-way Venn-filtration protocol between exosomal proteins from the SpC Venn filtration and the DPP data sets identified 13 HIV-BE and 34 HIV-SE high-confidence proteins, defined as DPP that exist in every exosomes sample and absent in every exo-free sample (Fig. 5e, supplemental Table 4). Similarly, 14 HIV+BE and 57 HIV+SE high-confidence DPP were identified (Fig. 6e, supplemental Table 4). This systematic analysis reveals previously unknown complex donor, tissue, compartment, and disease dependent variabilities in blood and semen exosome proteomes.

Identification of blood and semen exosomes markers modulated by HIV: To identify HIV-induced perturbations in blood and semen exProteins without bias based on known functional attributes, we used the lists of proteins identified by the SpC method in Figs. 5b and 6b. No common proteins between BE and SE were identified whether in HIV- (Fig. 7a) or HIV+ donors (Fig. 7b) despite the presence of 54 of the Top-100 exosomal markers in the lists (Fig. 2b, Supplemental Fig. 1). Furthermore, western blot validation did not identify a universal exosomal
marker that is only present in exosomes and absent in exosome-free plasma (Supplemental Fig. 6). The only exception was CD63 which was identified by western blot but absent in our stringent proteomics list. However, when analytical parameter were relaxed (SpC_851) CD63 was present with one peptide. One explanation for this discrepancy may be due to the differences in the levels of CD63 and in assay sensitivity. These results indicate that semen and blood, although autologous, are two compositionally different body-fluids.

Furthermore, HIV- and HIV+ sub-proteomes are also variable. While there are 6 and 8 proteins unique to BE of HIV- and HIV+ donors, respectively (Fig. 7c), 22 and 43 proteins are unique to SE from HIV- and HIV+ donors (Fig. 7d). Among the identified proteins, 8 are common in BE (Fig. 7c) and 16 are common in SE (Fig. 7d), irrespective of donor HIV status. Notably, SE contains more proteins than BE at each level of Venn-filtration and irrespective of HIV infection (Figs. 7a-d). On the other hand, enrichment scores of proteins perturbed by HIV using the SpC_Adjusted_AUC list, identified 28 (ADIPOQ, C6, CPB2, EFEMP1, FCGBP, FCN3, FN1, HBA1, IGHD, IGHV3-53, IGHV4-28, IGHV7-4-1, IGKV1-27, IGLC6, IGLV1-40, IGLV2-14, IGLV4-69, KNG1, MASP1, NEO1, PIP, SAA4, SEMG1, SERPINF1, VCL, VCP, VWF and YWHAZ) and 12 (NGAL (LCN2), FGG, ALDR, OS9, SBSPO, TIG1 (RARRES1), APOA1, FGA, ANXA4, CCT6A, AATC, and MUC5B) DPP in blood (Figs. 7e-f) and semen (Fig. 7g-h) exosomes respectively. The number of HIV-induced DPPs in BE is more than double that of SE, even though proteins identified in the semen proteome is larger than in the blood proteome. This observation points to the possibility that HIV-induced perturbation may be less frequent in semen compared to blood.

Gene ontology (GO) and KEGG pathway enrichment analyses revealed that BE proteins are enriched in processes related to protein activation cascade and humoral immune response mediated by FCN3, KNG1, MASP1, VWF, and SEMG1, as well as complement and coagulation cascades mediated by KNG1, VWF (Table 6). In contrast, SE proteins are enriched in processes that include regulation of heterotypic adhesion and positive regulation of substrate adhesion-
dependent cell spreading mediated by FGA, FGG, and APOA1; lipase inhibitor activity mediated by OS9 and ANXA4, complement and coagulation cascades mediated by FGA and FGG, and IL-17 signaling pathway by LCN2 and MUC5B (Table 7). Although both BE and SE KEGG pathways are enriched in complement and coagulation cascades, the participating proteins for BE and SE are different. In BE were FCN3, KNG1, MASP1, VWF, and SEMG1, while SE contained fibrinogen (FGA and FGG), a glycoprotein essential for clot formation. Whether SE-associated fibrinogen plays a role in seminal clotting and in liquefaction through "fibrinolytic" activities remains to be determined. However, fibrinolytic factors, including tissue plasminogen (t-PA), urinary plasminogen activator (u-PA), plasmin, fibrinogen, plasminogen, plasminogen activator inhibitor-1 (PAI-1), factor VIII coagulant activity (VIII:c) and fibrin monomers are present in semen [68]. Since SE DDPs (ANXA4 and CCT6A) are highly enriched in HIV+ group and ANXA4 and CCT6A are known to have paralogs, we compared ANXA4 and CCT6A family members in BE and SE. The analysis reveal that SE are significantly enriched in ANXA4 and CCT6A paralogs compared to BE (Fig 7i) regardless of clinical group. Within the clinical groups, SE from HIV+ group contained more of the ANXA4 and CCT6A paralogs compared to SE from HIV- group (Fig 7i). This enrichment of ANXA4 and CCT6A was validated by Western blot analysis (Fig. 7j). Whether or not SE -ANXA4 and -CCT6A paralogs interact with one another or promote physiological or pathophysiological functions is yet to be determined. PPI analysis (Molecular Interaction Search Tool, MIST [44], https://fgrtools.hms.harvard.edu/MIST/, October 16, 2019) revealed that ANXA4 and CCT6A may interact through MAPT, that promotes microtubule assembly and stability, CDH1, an epithelial cadherin involved in the formation of adherens junctions, AGR2, known for the production of the intestinal mucin, and EZH2, an important histone methyltransferase known to silence HIV gene expression [69] (Fig. 7k). The PPI is predicted to result in macromolecular complexes, including transcription factor complex & regulation of cyclin-dependent protein kinase activity, transcriptional repressor complex, negative regulation of transcription, DNA-dependent, protein polymerization & gamma-tubulin complex, regulation of actin filament polymerization,
ITGA4-ITGB1-VCAM1 complex, TNF-alpha/NF-kappa B signaling complex, histone deacetylase complex, ubiquitin ligase complex, SWI/SNF complex & nucleosome disassembly, E2F-6 complex, and negative regulation of cyclin-dependent protein kinase activity (Fig. 7k).

Mapping DPP•HIV interaction: Some of the GO Terms and interactome networks identified in our analyses intersect processes modulated by HIV and provide a framework for interpreting both previously identified and novel DPP from HIV+ BE and SE. Thus, we compared our DPP to other HIV-related datasets, including previously published HIV•Human PPIs and host factors implicated in HIV infection from the HIV•Human Interaction database [70]. We generated DPP•HIV interactome with ANXA4 and CCT6A because they are abundantly enriched in HIV+ group. The result indicates that ANXA4 (Fig. 8a) and CCT6A (Fig. 8b) intersect with HIV proteins. In this regard, the identified interactions include linkage of the DPP with HIV proteins (Figs. 8a,b). Given the presence of multiple ANXA and CCT paralogs in our dataset, and the fact that paralogs are likely to participate in the same molecular and/or genetic functions as their partners, we included paralogous pairs (gray circles) in the DPP•HIV interaction analysis. This analysis expanded the list of SE DPP that intersect HIV-modulated process (Fig. 8a,b, gray circles). Both ANXA4 and CCT6A and their paralogs have multiple interactions with HIV proteins; including the ability of gp120 to upregulate ANXA4 [71], Tat to upregulate CCT6A [72] and Vif to interact with CCT6A [73]. Interaction studies between SE and HIV proteins was not possible because both BE and SE likely do not contain viral particles as suggested by the non-detectable viral loads (<20 viral RNA copies/ml of blood) in the blood (Supplemental Table 1), and by the inability of these exosomes to significantly activate HIV LTR (Fig. 8c) or infect target cells (Fig. 8d). The absence of viral particles or various viral proteins may be due to the fact that the exosomes were isolated from donors with effective combination antiretroviral therapy (cART)-mediated HIV suppression (Supplemental Table 1).
Semen but not blood exosomes promote Protein•Nucleic acid binding activity independent of HIV: While attention has focused on cellular PPI in the formation of functional macromolecules, SE contain factors that have high affinity for molecular interactions (Welch et al., 2018). We therefore hypothesized that because SE is enriched with proteins that play roles in binding activities, extracts from both HIV- and HIV+ SE but not BE will bind host transcription factor sequences as previously described (Welch et al., 2018). Indeed, EMSA assays revealed that BE extracts did not bind p65 sequences (Fig. 9a, lanes 3 and 4), whereas SE extracts resulted in a clear shift of NF-kB p65 complex (Fig. 9a, lanes 5 and 6). Addition of 50X (Fig. 9b, lane 4) and 100X (Fig. 9b, lane 5) of cold NF-kB probe showed decreasing pattern of NF-kB shift (Fig. 9b, lanes 4-5). Analysis of EFBPs (HIV- and HIV+) and EFSPs (HIV- and HIV+) showed absence of DNA binding activity in the exosome-free fractions, pointing to compartmentalization of the DNA binding proteins in SE (Supplemental Fig. 7). Further analysis showed affinity of SE-containing DNA binding factors for other promoter sequences as evidenced by the formation of SE•Sp1, SE•AP1, and SE•IL1β complexes (Fig. 9b, lanes 1 to 3, respectively). Finally, SE also bound random single-stranded DNA sequences (Fig. 9b lane 6). Of note, HIV did not impair the DNA binding ability of SE. Together, these data indicate that SE but not BE contain DNA binding factors that may regulate host and HIV transcription as previously reported (Welch et al., 2018).

HIV diminished the ability of SE to function as cell adhesion substrate for U937 monocytic and Jurkat T cells: Probing further into the binding properties of BE and SE, we sought to determine if BE- or SE- proteins are involved in functional PPIs. In static cell clustering and cell adhesion assays using U937 monocytic and Jurkat T cell lines seeded atop vehicle-, BE-, or SE-coated wells, BE did not alter the clustering ability or lack thereof for U937 monocytes and Jurkat T cells when compared to vehicle controls. However, Jurkat cells seeded on SE-coated wells were more dispersed, while no effect of SE was observed on U937 cells (Fig. 10a). Following removal of non-adherent cells, the adhesive interaction between U937 or T cells and BE was
similar to the interaction with vehicle-treated cells (Figs. 10b-d). In contrast, adhesive interaction of monocytes and T cells with SE was enhanced in comparison to vehicle or BE treatment (Figs. 10b-d). Interestingly, SE from HIV+ donors were less efficient in mediating monocyte adhesion, an observation that was also made for Jurkat T cells, although to a lesser extent (Figs. 10b-d). On the other hand, cell clustering mediated by EFBP and EFSP (Supplemental Fig. 8a) are similar to the phenotype observed in BE and vehicle (Fig. 10a). Moreover, EFBP and EFSP regardless of HIV infection were not as efficient as SE in promoting cell adhesion (Supplemental Fig. 8b-d). The adhesive factors are enriched in SE since the differential adhesive ability of SE is significantly higher than that of EFSP, regardless of HIV status (Supplemental Fig. 8e). The subtle background adhesion observed in vehicle control may reflect presence of adhesion molecules recruited from the serum-containing medium, since experiments were performed in the presence of serum. These functional analyses validated the GO and KEGG predicted functions of SE (Table 7) and reveal that HIV may impair the ability of SE to bind proteins on the cell surface. Indeed we have recently shown that monocytes pre-treated with SE from healthy or ART-naive HIV+ donors significantly affected the actin cytoskeleton reorganization of monocytes and promoted their adhesion atop of collagen [74].

Discussion

All biological processes are driven by complex distal and/or proximal interactions between cellular and extracellular biomolecules, such as those present in exosomes. To this end, Omics studies especially transcriptomics and proteomics are leveraged for multifaceted analysis of biomolecules in host cells/tissues and their extracellular partners for various purposes. However, the majority of Omics studies depend on pooled samples analyzed in replicates. While this approach has some advantages, it overlooks inherent inter-donor variabilities in clinical or in vivo-derived samples. Variations in donor samples are particularly important in assessing the cargo of
exosomes where subtle differences in the levels of individual features within exosome secreting cells may or may not be captured as exosomal cargo. Given that inter-donor variation exist even between genetically identical twins and is driven mostly by difficult to control extrinsic environmental factors [75], we adopted the SMPP pipeline described in this study for comprehensive analysis of exosome cargo composition taking into account donor variability. SMPP is especially relevant to MS proteomic methods that utilize peptide quantities and features in the calculation of protein fold change or abundance. The extent to which donor-dependent variations influence the presence and abundance of exosomal cargo, such as those of BE vs SE proteins is biologically relevant. Not only will donor variabilities affect analytical output, it will also affect interpretation of cellular and/or viral processes that the proteins regulate. SMPP pipeline accounts for donor variabilities, compensates for other sources of variation/noise (sample purification, peptide behavior), while also quantifying relative protein abundances for each protein. Thus, SMPP ensures that, for any identified protein, its presence and enrichment or lack thereof could not be the result of significantly over/under abundance in a few individuals which could have skewed the pooled sample. We used SMPP to provide a comprehensive, unbiased profile of exProteins in autologous BE and SE of HIV- and HIV+ donors. We overcame the complexity of analyzing exosome cargo by performing our analysis on paired exosome-free fractions to help with identification of true exosome-associated proteins. Our findings were validated using unpaired pooled BE and SE samples. This approach provides a snapshot of exProteins present in autologous blood and semen fractions, identifies donor variabilities, and captures body-fluid and compartment protein distributions.

Despite the presence of donor- and body-fluid- dependent variations in protein content and compartmentalization in blood and seminal plasma, common proteins were identified in BE, EFBP, SE, and EFSP. Noteworthy is that a head to head comparison of proteins from the two body-fluid types (blood vs semen) reveals that the semen proteome is more complex than the autologous blood proteome (Figs. 2c,d; 4a; 5a; and 6a). Comparison of BE and SE proteomes
from other studies [76-78] with our dataset provides a list of protein overlaps and differences (Fig. 2b, Supplemental Figs. 2-3). Although the number of identified proteins in our study is relatively smaller compared to other studies, it is noteworthy that our study analyzed individual blood and semen proteomes in a donor-dependent manner as opposed to the pooled analysis used in other studies to generate aggregate data [76-78]. Individualized analysis of autologous blood and semen while reducing the total concentration of proteins, facilitates identification of the degree of BE and SE protein cargo variation within the same donors. Of interest, the degree of overlap or difference is significant, underscoring the need for autologous single-donor analysis when possible; to reduce overestimation of results.

The use of autologous specimens under physiological and pathophysiological conditions allowed us to uncover proteins that are common to blood and semen. For example, our results indicate that components of humoral immunity are encapsulated in BE and SE. The humoral immunity proteins—IGHG2 and IGKC have the potential to serve as a body-fluid protein biosignature. IGKC is a product of plasma cells and a B-cell signature shown to be present in blood and tumors. It is not surprising that humoral immunity factors are present in blood but whether the presence of IGKC in seminal fractions is directly related to increased numbers of plasma cells secreting Ig in semen is unknown. IGKC protein however has been found in semen [79] and it is a single immune marker that allows prediction of metastasis-free survival and response to chemotherapy [55, 80]. Since IGKC is already used as an immunologic biomarker of prognosis and response to therapy in solid tumors, our finding opens the way for a broader use of IGKC as a marker of global exProteins.

We also identified body-fluid-specific enrichment of proteins and pathways that are involved in reproduction and HIV pathogenesis. One such protein is CLU, an extracellular chaperon glycoprotein, which is significantly enriched in SE, irrespective of HIV infection (Fig. 4c-f). CLU concentration in semen (0.4 -15.0 mg/ml) is about 20-fold higher than that in blood [56, 57] and this observation was supported by our data (Fig. 4f). The high levels of CLU in semen
suggests a potential role in reproduction and/or transmission of sexually transmitted agents. Indeed, seminal CLU has been implicated in many reproductive processes, including female tolerogenic response to male antigens, spermatozoa maturation, lipid transport, and sperm membrane remodeling [81]. With its putative link to adhesion molecules, such as fibronectin, CLU may indirectly promote cell adhesion. As shown in Fig. 10a-c, monocytes and T cell adhesions to SE as substrate were increased compared to vehicle control and BE, an indication that SE contain protein binding factors that mediate cellular adhesion. This finding is surprising because similar to SE, BE are enriched in immunoglobulins and other adhesive factors. Despite the ability of SE-containing factors to promote cell adhesion, HIV infection diminished the adhesive ability of SE suggesting a potentially novel but yet obscure mechanism by which the virus modulates host cellular pathways. Given that the SE-associated proteins that regulate cell adhesion is yet to be identified, future studies are essential to understand the functional relevance of this finding.

Moreover, seminal CLU, but not blood CLU is composed of a set of fucosylated glycans with high affinity for DC-SIGN [63]. This property of CLU is important in HIV pathogenesis since CLU glycan motifs mediate the effective binding of seminal CLU to DC-SIGN, which efficiently blocks the binding of HIV to DCs. Additionally, seminal CLU promotes the uptake of stress-damaged proteins by dendritic cells via DC-SIGN [82]. Given the importance of CLU in semen biology both in reproduction and infectious diseases, we showed that seminal CLU is mostly enriched in SE compared to EFSP (Fig. 4f). This is a significant observation because SE may function to provide protection to CLU from degradation and serve as vehicle for CLU transport to distal sites, including mucosal sites, where CLU is needed for reproductive purposes or blockade of DC-SIGN-mediated HIV infection. Indeed, we have previously reported that SE potently inhibit HIV infection through multiple mechanisms [5, 11, 19-21], but the anti-HIV agent in SE is unknown. Since semen blocks the recognition of HIV by DC-SIGN [64, 83] and inhibits DC-SIGN-mediated transmission of HIV [83], it is necessary to evaluate the role of SE-CLU in SE-mediated HIV inhibition.
Since most proteins fulfill their functional potential in multi-protein complexes, the presence of paralogous ANXAs and CCTs proteins in semen suggest that such supramolecular assemblages may partake in molecular interactions with viral proteins or host proteins involved in reproduction. An example is the enrichment of CCT family of proteins in semen (Fig. 7i,j). In somatic cells, CCTs function as molecular chaperones that assist in the correct folding of cellular proteins [84], including the folding of cytoskeletal proteins (actin and tubulin). Although the role of CCT family of proteins in reproduction is not known, the initial identification of TCP-1 complex was as CCT1—a highly expressed mouse testicular protein. In the testes, CCT complex are implicated in reorganization of microtubular cytoskeleton that occurs during differentiation of spermatids to spermatozoa [85]. CCT complex is also implicated in capacitation-dependent binding of spermatozoa to zonae pellucidae [86]. As with testicular CCTs, SE are enriched in CCTs. Whether or not SE-associated CCT family of proteins are active, cytoskeletal proteins bind CCTs in a sequence-specific manner, and it is possible that other host and viral proteins could use CCTs as a general binding interface. Therefore, SE-CCT may be linked to many viral and cellular processes that depend on cytoskeletal system integrity and formation of multiprotein complexes. Similar to CCT-mediated regulation of reproductive processes via binding of spermatozoa to zonae pellucidae [86], ANXA4 binds membrane phospholipids to regulate many biological processes, including cell adhesion, negative regulation of NF-kB activity, negative regulation of interleukin-8 secretion, and regulation of transcription by RNA polymerase II [87, 88].

An earlier study showed that HIV infection alters the level and type of proteins found in different extracellular vesicles [22], which may result in positive or negative effect on HIV pathogenesis. PPI networks and modulations of such networks by viral proteins may be one of the important mechanisms of local interactions in BE and SE in infected tissues and within body-fluids. Such interactions may contribute to the protein compartmentalization pattern observed between HIV- and HIV+ groups. Our results suggest that HIV infection modulated enrichment pattern of host proteins in exosomes, and such enrichment may be body-fluid-dependent. For
example, HIV infection results in significant enrichment of ANXA4 and CCT6A proteins in SE but not in BE. EMSA assays indicate that SE but not BE contain proteins that bind host transcription factors and DNA sequences, as well as proteins that promote cell adherence. These properties of SE are absent in their autologous BE—underscoring the body-fluid differences in function.

While the significance of protein paralogy is unknown, it is possible that SE-containing ANXA4 and CCT6A paralogs may participate in the same molecular and functional interactions as their paralogs. SE-enriched ANXA4, CCT6A and their paralogs may regulate the assemblage of multiprotein complexes that may include, nucleic acid binding and adhesion molecules. While SE contain concentrations of adhesion molecules sufficient to elicit monocyte and T cell adhesions, it is yet to be determined what effect, if any, SE will have on aggregation of sperm cells, especially since spermatozoa in the epididymal lumen are dispersed. This observation leads to more fundamental questions concerning the balance of factors that promote or prevent aggregation in SE. It is possible that in vivo, other components that function as anti-adhesion factors may be present to modulate the function of adhesion molecules. Indeed, anti-sticking factor (ASF-I) that functions to inhibit adhesion of spermatozoa to substrates has been characterized [89]. Based on this analysis, we conclude that the literature curated function of SE proteins may not be due to simple interactions among paralogs but a reflection on the formation of larger complexes by functionally related proteins or Protein•Nucleic acid complexes, such as observed in this study (Fig. 9). Other proteins including aldose reductase (AKR1B1), SBSPON, and OS9 are found to be enriched in HIV- SE. AKR1B1 has paralogous proteins although the paralogs are absent in SE.

The SMPP pipeline facilitated identification of HIV-induced perturbations in blood and semen proteomes, and how these perturbations are compartmentalized in exosomes and their exosomes-free plasma. The mechanisms or determinants of protein compartmentalization in exosomes or exosome-free fractions of blood and semen in general is unknown and the relevance of protein compartmentalization with regards to function remains to be determined. However, local
protein interactions in the context of tissue compartments may be important for reproduction and HIV pathogenesis. A previous study demonstrated protein compartmentalization in a sub-group of seminal vesicles—epididymosomes that are secreted by epididymal epithelial cells [90]. In that regard, epididymosomes have been shown to coordinate the association of epididymal proteins, such as ganglioside GM1 and sphingomyelin [91]. Whether or not protein compartmentalization in BE and SE observed in this study would influence compartmentalization of HIV particles between semen and blood as reported previously [92] is unknown. We anticipate that our data would provide important pathophysiological insights into the role of BE and SE in HIV pathogenesis.

In this study, we did not find common proteins between BE and SE in HIV- or HIV+ groups (Fig. 7a,b), despite the presence of 54 proteins of the top-100 most abundant exosomal proteins in our lists (Fig. 2b). This discrepancy may be because exosome cargo largely varies between donors, originating cells, and body-fluid (blood vs semen). Furthermore, the proteome of exosome-free fractions are not traditionally taken into account in most studies, and donor variability is also rarely accounted for. Our data show that the use of SMPP when possible may facilitate identification of specimen-specific exosomal biosignature.

We also did not find HIV proteins in any of the blood and seminal plasma fractions through proteomics, despite these proteins being reported as present in urine EVs [93]. Reasons for this discrepancy may include the level of viral suppression in study participants, the type of vesicles involved, and method of vesicle isolation/proteomic analysis. Our study participants have undetectable viral load and were on therapy (Supplemental Table 1). The absence of HIV proteins suggests that ART may have effectively reduced both blood and seminal viral loads to undetectable levels in all participants, as previously indicated [94]. It is worth noting that our study did not include viremic donors who were not on therapy. Thus, it is difficult to ascribe any observed differences between HIV- and HIV+ groups to only HIV or to the drugs used to treat HIV, especially because ART are associated with both BE and SE [21].
Although HIV proteins were not detected in our studies, nonetheless, the biological nature of the pathways and DPP•HIV interactome network enabled us to discuss potential impact of BE and SE proteins in physiological and pathophysiological conditions. Collectively, our prospective analysis of the body-fluid-specific exProtein signatures of blood and semen in physiological and pathophysiological states provides new insights on the biology and importance of exosomes in reproduction and HIV pathogenesis. In particular, we identified functional differences between BE and SE, as well as a panel of strong, independent BE and SE markers that may be developed into exosome-based predictive signatures of male fertility or HIV infection.

**Acknowledgements:** This work was supported by the National Institute on Drug Abuse (NIDA) grant 1R01DA042348-01 (to CMO), VA Merit Review BX000207 and National Institutes of Health (NIH) 5T32AI343 (to JTS) and NIH 5T32AI007533-18 (to JLW). CD is partly supported by NIH grants DA021471, Al22960, and Al110527. Both CD and SP are partly supported by NIH grants MD007586 and MD007593 (to SP). We thank the semen and blood donors for providing samples. We thank the University of Iowa Hospitals and Clinics Virology (HIV/AIDS) clinic, and Dr. Amy Sparks, Ryan Brumm, and Jennifer Jagnow of the University of Iowa Reproductive Testing Laboratories for assistance with specimen collection. We are also grateful to the Meharry Medical Center CORES (Proteomics & Bioinformatics). Our gratitude also goes to Dr. Bryson Okeoma of Stony Brook University for critical review of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Data Availability:** The authors declare that all data supporting the findings of this study are available within the article and its supplemental information files. Data are available via ProteomeXchange with identifier PXD013033. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Author contributions: CMO conceived studies. VP, JS, SP, CD, CMO designed the study. HK, YL, JLW, VP, JP, JS, SP, CD, CMO participated in data analysis. HK, YL, and JLW performed biological and validation experiments. VP acquired proteomics data. HK, VP, SP, SNM, and SSS, performed bioinformatics analysis. All authors participated in manuscript preparation and approved the final version of the manuscript.

Competing interests: The authors declare no competing interests.

Materials & Correspondence: Correspondence and material requests should be addressed to Dr. Chioma Okeoma at chioma.okeoma@stonybrook.edu
References

1. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C and Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808-15. doi: 10.1093/nar/gks1094
2. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447-52. doi: 10.1093/nar/gku1003
3. Raposo G and Stoorvogel W (2013) Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol 200:373-83. doi: 10.1083/jcb.201211138
4. Madison MN and Okeoma CM (2015) Exosomes: Implications in HIV-1 Pathogenesis. Viruses 7:4093-118. doi: 10.3390/v7072810
5. Madison MN, Roller RJ and Okeoma CM (2014) Human semen contains exosomes with potent anti-HIV-1 activity. Retrovirology 11:102. doi: 10.1186/s12977-014-0102-z
6. Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, Strobl J, Westerberg K, Gottardo R, Tewari M and Hladik F (2014) Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. Nucleic Acids Res 42:7290-304. doi: 10.1093/nar/gku347
7. Welch JL, Stapleton JT and Okeoma CM (2019) Vehicles of intercellular communication: exosomes and HIV-1. J Gen Virol. doi: 10.1099/jgv.0.001193
8. Shtam TA, Kovalev RA, Varfolomeeva EY, Makarov EM, Kil YV and Filatov MV (2013) Exosomes are natural carriers of exogenous siRNA to human cells in vitro. Cell Commun Signal 11:88. doi: 10.1186/1478-811x-11-88
9. Tomasoni S, Longaretti L, Rota C, Morigi M, Conti S, Gotti E, Capelli C, Introna M, Remuzzi G and Benigni A (2013) Transfer of growth factor receptor mRNA via exosomes unravels the regenerative effect of mesenchymal stem cells. Stem Cells Dev 22:772-80. doi: 10.1089/scd.2012.0266
10. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ and Lotvall JO (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol 9:654-9. doi: 10.1038/ncll1596
11. Madison MN, Jones PH and Okeoma CM (2015) Exosomes in human semen restrict HIV-1 transmission by vaginal cells and block intravaginal replication of LP-BM5 murine AIDS virus complex. Virology 482:189-201. doi: 10.1016/j.virol.2015.03.040
12. Robbins PD and Morelli AE (2014) Regulation of immune responses by extracellular vesicles. Nat Rev Immunol 14:195-208. doi: 10.1038/nri3622
13. Pinto-Neto LF, Vieira NF, Soprani M, Cunha CB, Dietze R and Ribeiro-Rodrigues R (2002) Lack of correlation between seminal and plasma HIV-1 viral loads is associated with CD4 T cell depletion in therapy-naive HIV-1+ patients. Mem Inst Oswaldo Cruz 97:563-7.
14. Chaix ML, Boufassa F, Meyzer C, Leruez-Ville M, Mahjoub N, Nere ML, Genet P, Duivivier C, Lascoux-Combes C, Lambotte O and Ghosn J (2017) Detectable HIV-RNA in semen of HIV controllers. PLoS One 12:e0183376. doi: 10.1371/journal.pone.0183376
15. Cheret A, Durier C, Melard A, Ploquin M, Heitzmann J, Lecouroux C, Avetteland-Fenoel V, David L, Pialoux G, Chennebault JM, Muller-Trutwin M, Goujard C, Rouzioux C and Meyer L (2017) Impact of early cART on HIV blood and semen compartments at the time of primary infection. PLoS One 12:e0180191. doi: 10.1371/journal.pone.0180191
16. Jiao YM, Chen GL, Zhu WJ, Huang HH, Fu JL, Chen WW, Shi M, Zhang T, Wu H and Wang FS (2017) Higher viral load and genetic diversity of HIV-1 in seminal compartments than in blood of seven Chinese men who have sex with men and have early HIV-1 infection. Microbiol Immunol 61:239-246. doi: 10.1111/1348-0421.12488
17. Osborne BJW, Marsh AK, Huibner S, Shahabi K, Liu C, Contente T, Nagelkerke NJD, Kovacs C, Benko E, Price L, MacDonald KS and Kaul R (2017) Clinical and Mucosal Immune Correlates of HIV-1 Semen Levels in Antiretroviral-Naive Men. Open Forum Infect Dis 4:ofx033. doi: 10.1093/ofid/ofx033

18. Pasquier C, Walschaerts M, Raymond S, Moinard N, Saune K, Daudin M, Izopet J and Bujan L (2017) Patterns of residual HIV-1 RNA shedding in the seminal plasma of patients on effective antiretroviral therapy. Basic Clin Androl 27:17. doi: 10.1186/s12610-017-0063-x

19. Welch JL, Kaddour H, Schlievert PM, Stapleton JT and Okeoma CM (2018) Semen exosomes promote transcriptional silencing of HIV-1 by disrupting NF-kB/Sp1/Tat circuitry. J Virol. doi: 10.1128/jvi.00731-18

20. Welch JL, Madison MN, Margolick JB, Galvin S, Gupta P, Martinez-Maza O, Dash C and Okeoma CM (2017) Effect of prolonged freezing of semen on exosome recovery and biologic activity. Sci Rep 7:45034. doi: 10.1038/srep45034

21. Welch JL, Kaddour H, Winchester L, Fletcher CV, Stapleton JT and Okeoma CM (2019) Semen extracellular vesicles from HIV-1 infected individuals inhibit HIV-1 replication in vitro, and extracellular vesicles carry antiretroviral drugs in vivo. JAIDS: Journal of Acquired Immune Deficiency Syndromes.

22. Anyanwu SI, Doherty A, Powell MD, Obialo C, Huang MB, Quarsheie A, Mitchell C, Bashir K and Newman GW (2018) Detection of HIV-1 and Human Proteins in Urinary Extracellular Vesicles from HIV+ Patients. Adv Virol 2018:7863412. doi: 10.1155/2018/7863412

23. Sinha A, Principe S, Alfaro J, Ignatenko A, Ignatenko V and Kislinger T (2018) Proteomic Profiling of Secreted Proteins, Exosomes, and Microvesicles in Cell Culture Conditioned Media. Methods Mol Biol 1722:91-102. doi: 10.1007/978-1-4939-7553-2_6

24. Bakhtyar N, Jeschke MG, Herer E, Sheikholeslam M and Amini-Nik S (2018) Exosomes from acellular Wharton's jelly of the human umbilical cord promotes skin wound healing. Stem Cell Res Ther 9:193. doi: 10.1186/s13287-018-0921-2

25. Diaz-Varela M, de Menezes-Neto A, Perez-Zsolt D, Gamez-Valero A, Segui-Barber J, Izquierdo-Useros N, Martinez-Picado J, Fernandez-Becerra C and Del Portillo HA (2018) Proteomics study of human cord blood reticulocyte-derived exosomes. Sci Rep 8:14046. doi: 10.1038/s41598-018-32386-2

26. Eirin A, Zhu XY, Woollard JR, Tang H, Dasari S, Lerman A and Lerman LO (2019) Metabolic Syndrome Interferes with Packaging of Proteins within Porcine Mesenchymal Stem Cell-Derived Extracellular Vesicles. Stem Cells Transl Med. doi: 10.1002/sctm.18-0171

27. Gualdron-Lopez M, Flannery EL, Kangwanransan N, Chuenchob V, Fernandez-Orth D, Segui-Barber J, Royo F, Falcon-Perez JM, Fernandez-Becerra C, Lacerda MVG, Kappe SHI, Sattabongkot J, Gonzalez JR, Mikolajczak SA and Del Portillo HA (2018) Characterization of Plasmodium vivax Proteins in Plasma-Derived Exosomes From Malaria-Infected Liver-Chimeric Humanized Mice. Front Microbiol 9:1271. doi: 10.3389/fmicb.2018.01271

28. Jimenez L, Yu H, McKenzie A, Franklin JL, Patton JG, Liu Q and Weaver AM (2019) Quantitative proteomic analysis of small and large extravasal extracellular vesicles (EVs) reveals enrichment of adhesion proteins in small EVs. J Proteome Res. doi: 10.1021/acs.jproteome.8b00647

29. Kimura Y, Kojima M, Kurosawa T, Sasaki R, Ichihara S, Hiraku Y, Tomimoto H, Murata M and Oikawa S (2018) Proteomic Profiling of Exosomal Proteins for Blood-based Biomarkers in Parkinson's Disease. Neuroscience 392:121-128. doi: 10.1016/j.neuroscience.2018.09.017

30. Kuang M, Tao X, Peng Y, Zhang W, Pan Y, Cheng L, Yuan C, Zhao Y, Mao H, Zhuge L, Zhou Z, Chen H and Sun Y (2019) Proteomic analysis of plasma exosomes to differentiate malignant from benign pulmonary nodules. Clin Proteomics 16:5. doi: 10.1186/s12014-019-9225-5

31. Luo D, Zhan S, Xia W, Huang L, Ge W and Wang T (2018) Proteomics study of serum exosomes from papillary thyroid cancer patients. Endocr Relat Cancer 25:879-891. doi: 10.1530/erc-17-0547
32. Shushkova NA, Vavilov NE, Novikova SE, Farafonova TE, Tikhonova OV, Liao PC and Zgoda VG (2018) [Quantitative proteomics of human blood exosomes]. Biomed Khim 64:496-504. doi: 10.18097/pbmc20186406496
33. Whitham M and Febbraio MA (2019) Redefining Tissue Crosstalk via Shotgun Proteomic Analyses of Plasma Extracellular Vesicles. Proteomics 19:e1800154. doi: 10.1002/pmic.201800154
34. Zhang H, Liu J, Qu D, Wang L, Wong CM, Lau CW, Huang Y, Wang YF, Huang H, Xia Y, Xiang L, Cai Z, Liu P, Wei Y, Yao X, Ma RCW and Huang Y (2018) Serum exosomes mediate delivery of arginine 1 as a novel mechanism for endothelial dysfunction in diabetes. Proc Natl Acad Sci U S A 115:E6927-e6936. doi: 10.1073/pnas.1721521115
35. Kodidela S, Wang Y, Patters BJ, Gong Y, Sinha N, Ranjit S, Gerth K, Haque S, Cory T, McArthur C, Kumar A, Wan JY and Kumar S (2019) Proteomic Profiling of Exosomes Derived from Plasma of HIV-Infected Alcohol Drinkers and Cigarette Smokers. J Neuroimmune Pharmacol. doi: 10.1007/s11481-019-09853-2
36. Rydze RT, Bhattarai N and Stapleton JT (2012) GB virus C infection is associated with a reduced rate of reactivation of latent HIV and protection against activation-induced T-cell death. Antivir Ther 17:1271-9. doi: 10.3851/imp2309
37. Madison MN, Welch JL and Okeoma CM (2017) Isolation of Exosomes from Semen for in vitro Uptake and HIV-1 Infection Assays. Bio Protoc 7. doi: 10.21769/BioProtoc.2216
38. Momen-Heravi F, Bala S, Bukong T and Szabo G (2014) Exosome-mediated delivery of functionally active miRNA-155 inhibitor to macrophages. Nanomedicine: Nanotechnology, Biology and Medicine 10:1517-1527. doi: https://doi.org/10.1016/j.nano.2014.03.014
39. Lin H, He L and Ma B (2013) A combinatorial approach to the peptide feature matching problem for label-free quantification. Bioinformatics 29:1768-75. doi: 10.1093/bioinformatics/btt274
40. Consortium TU (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47:D506-D515. doi: 10.1093/nar/gky1049
41. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA and Ma B (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 11:M111.010587. doi: 10.1074/mcp.M111.010587
42. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM and Yates JR, 3rd (1999) Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 17:676-82. doi: 10.1038/10890
43. Wang J, Vasaikar S, Shi Z, Greer M and Zhang B (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic acids research 45:W130-W137.
44. Hu Y, Vinayagam A, Nand A, Comjean A, Chung V, Hao T, Mohr SE and Perrimon N (2018) Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. Nucleic Acids Res 46:D567-d574. doi: 10.1093/nar/gkx1116
45. Marcu A, Arndt D, Grant JR, Babicki S, Liang Y, Wishart DS and Maciejewski A (2016) Heatmapper: web-enabled heat mapping for all. Nucleic Acids Research 44:W147-W153. doi: 10.1093/bioinformatics/btw706
46. Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html.
47. Okeoma CM, Huegel AL, Lingappa J, Feldman MD and Ross SR (2010) APOBEC3 proteins expressed in mammary epithelial cells are packaged into retroviruses and can restrict transmission of milk-borne virions. Cell Host Microbe 8:534-43. doi: 10.1016/j.chom.2010.11.003
48. Mousseau G, Kessing CF, Fromentin R, Trautmann L, Chomont N and Valente ST (2015) The Tat Inhibitor Didehydro-Cortistatin A Prevents HIV-1 Reactivation from Latency. MBio 6:e00465. doi: 10.1128/mBio.00465-15
49. Tran NH, Qiao R, Xin L, Chen X, Liu C, Zhang X, Shan B, Ghodsi A and Li M (2019) Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. Nat Methods 16:63-66. doi: 10.1038/s41592-018-0260-3
50. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, Samuel M, Pathan M, Jois M and Chilamkurti N (2016) ExoCarta: a web-based compendium of exosomal cargo. Journal of molecular biology 428:688-692.
51. Lundgren DH, Hwang S-I, Wu L and Han DK (2010) Role of spectral counting in quantitative proteomics. Expert review of proteomics 7:39-53.
52. Megger DA, Bracht T, Meyer HE and Sitek B (2013) Label-free quantification in clinical proteomics. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1834:1581-1590.
53. Zhang Y, Wen Z, Washburn MP and Florens L (2015) Improving label-free quantitative proteomics strategies by distributing shared peptides and stabilizing variance. Analytical chemistry 87:4749-4756.
54. Giavarina D (2015) Understanding bland altman analysis. Biochemia medica: Biochemia medicina 25:141-151.
55. Schmidt M, Micke P, Gehrmann M and Hengstler JG (2012) Immunoglobulin kappa chain as an immunologic biomarker of prognosis and chemotherapy response in solid tumors. Oncoimmunology 1:1156-1158. doi: 10.4161/onci.21653
56. Choi NH, Tobe T, Hara K, Yoshida H and Tomita M (1990) Sandwich ELISA assay for quantitative measurement of SP-40,40 in seminal plasma and serum. J Immunol Methods 131:159-63.
57. O’Bryan MK, Baker HW, Saunders JR, Kirszbaum L, Walker ID, Hudson P, Liu DY, Glew MD, d’Apice AJ and Murphy BF (1990) Human seminal clusterin (SP-40,40). Isolation and characterization. J Clin Invest 85:1477-86. doi: 10.1172/jci114594
58. Sahlen G, Nilsson O, Larsson A, Carlsson L, Norlen BJ and Ronquist G (2010) Secretions from seminal vesicles lack characteristic markers for prostasomes. Ups J Med Sci 115:107-12. doi: 10.3109/03009730903366067
59. Reyes-Moreno C, Boilard M, Sullivan R and Sirard MA (2002) Characterization and identification of epididymal factors that protect ejaculated bovine sperm during in vitro storage. Biol Reprod 66:159-66.
60. Ibrahim NM, Gilbert GR, Loseth KJ and Crabo BG (2000) Correlation between clusterin-positive spermatozoa determined by flow cytometry in bull semen and fertility. J Androl 21:887-94.
61. O’Bryan MK, Murphy BF, Liu DY, Clarke GN and Baker HW (1994) The use of anticlusterin monoclonal antibodies for the combined assessment of human sperm morphology and acrosome integrity. Hum Reprod 9:1490-6.
62. Jenne DE and Tschopp J (1989) Molecular structure and functional characterization of a human complement cytolysis inhibitor found in blood and seminal plasma: identity to sulfated glycoprotein 2, a constituent of rat testis fluid. Proc Natl Acad Sci U S A 86:7123-7.
63. Sabatte J, Faigle W, Ceballos A, Morelle W, Rodríguez Rodríguez C, Remes Lenicov F, Thepaut M, Fieschi F, Malchiiodi E, Fernandez M, Arenzana-Seisdedos F, Lortat-Jacob H, Michalski JC, Geffner J and Amigorena S (2011) Semen clusterin is a novel DC-SIGN ligand. J Immunol 187:5299-309. doi: 10.4049/jimmunol.1101889
64. Sabatte J, Ceballos A, Raiden S, Vermeulen M, Nahmod K, Maggini J, Salamone G, Salomon H, Amigorena S and Geffner J (2007) Human seminal plasma abrogates the capture and transmission of human immunodeficiency virus type 1 to CD4+ T cells mediated by DC-SIGN. J Virol 81:13723-34. doi: 10.1128/jvi.01079-07
65. Ibrahim NM, Foster DN and Crabo BG (2001) Localization of clusterin on freeze-preserved bull spermatozoa before and after glass wool-sephadex filtration. J Androl 22:891-902.
66. Fukuda T, Miyake H, Enatsu N, Matsushita K and Fujisawa M (2016) Seminal level of clusterin in infertile men as a significant biomarker reflecting spermatogenesis. Andrologia 48:1188-1194. doi: 10.1111/and.12558
67. Liu H, Sadygov RG and Yates JR, 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76:4193-201. doi: 10.1021/ac0498563
68. Van Dreden P, Gonzales J and Poirot C (1991) Human seminal fibrinolytic activity: specific determinations of tissue plasminogen activator and urokinase. Andrologia 23:29-33.
69. Friedman J, Cho WK, Chu CK, Keedy KS, Archim NM, Margolis DM and Karn J (2011) Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. J Virol 85:9078-89. doi: 10.1128/jvi.00836-11
70. Ako-Adjei D, Fu W, Wallin C, Katz KS, Song G, Darji D, Brister JR, Ptak RG and Pruitt KD (2014) HIV-1, human interaction database: current status and new features. Nucleic acids research 43:D566-D570.
71. Molina L, Grimaldi M, Robert-Hebmann V, Espert L, Varbanov M, Devaux C, Granier C and Biard-Piechaczyk M (2007) Proteomic analysis of the cellular responses induced in uninfected immune cells by cell-expressed X4 HIV-1 envelope. Proteomics 7:3116-30. doi: 10.1002/pmic.200700306
72. Jarboui MA, Bidoia C, Woods E, Roe B, Wynne K, Elia G, Hall WW and Gautier VW (2012) Nucleolar protein trafficking in response to HIV-1 Tat: rewiring the nucleolus. PLoS One 7:e48702. doi: 10.1371/journal.pone.0048702
73. Luo Y, Jacobs EY, Greco TM, Mohammed KD, Tong T, Keegan S, Binley JM, Cristea IM, Fenyo D, Rout MP, Chait BT and Muesing MA (2016) HIV-host interactor complex revealed directly from infected cells. Nat Microbiol 1:16068. doi: 10.1038/nmicrobiol.2016.68
74. Lyu Y, Kaddour H, Kopcho S, Panzner TD, Shouman N, Kim E-Y, Martinson J, McKay H, Martinez-Maza O, Margolick JB, Stapleton JT and Okeoma CM (2019) Human Immunodeficiency Virus (HIV) Infection and Use of Illicit Substances Promote Secretion of Semen Exosomes that Enhance Monocyte Adhesion and Induce Actin Reorganization and Chemotactic Migration. Cells 8:1027.
75. Brodin P, Jojic V, Gao T, Bhattacharya S, Angel CJ, Furman D, Shen-Orr S, Dekker CL, Swan GE, Butte AJ, Maecker HT and Davis MM (2015) Variation in the human immune system is largely driven by non-heritable influences. Cell 160:37-47. doi: 10.1016/j.cell.2014.12.020
76. Poliakov A, Spilman M, Dokland T, Amling CL and Mobley JA (2009) Structural heterogeneity and protein composition of exosome-like vesicles (prostasomes) in human semen. The Prostate 69:159-167.
77. Yang C, Guo Wb, Zhang Wu, Bian J, Yang Jk, Zhou Qz, Chen Mk, Peng W, Qi T and Wang Cy (2017) Comprehensive proteomics analysis of exosomes derived from human seminal plasma. Andrology 5:1007-1015.
78. An M, Wu J, Zhu J and Lubman DM (2018) Comparison of an Optimized Ultracentrifugation Method versus Size-Exclusion Chromatography for Isolation of Exosomes from Human Serum. Journal of proteome research 17:3599-3605.
79. da Silva BF, Souza GH, Io Turco EG, Del Giudice PT, Soler TB, Spaine DM, Borrelli Junior M, Gozzo FC, Pilau EJ, Garcia JS, Ferreira CR, Eberlin MN and Bertolla RP (2013) Differential seminal plasma proteome according to semen retrieval in men with spinal cord injury. Fertil Steril 100:959-69. doi: 10.1016/j.fertnstert.2013.06.009
80. Whiteside TL and Ferrone S (2012) For breast cancer prognosis, immunoglobulin kappa chain surfaces to the top. Clin Cancer Res 18:2417-9. doi: 10.1158/1078-0432.crr-12-0566
81. Humphreys DT, Carver JA, Easterbrook-Smith SB and Wilson MR (1999) Clusterin has chaperone-like activity similar to that of small heat shock proteins. J Biol Chem 274:6875-81.
82. Merlotti A, Dantas E, Remes Lenicov F, Ceballos A, Jancic C, Varese A, Rubione J, Stover S, Geffner J and Sabatte J (2015) Fucosylated clusterin in semen promotes the uptake of stress-
damaged proteins by dendritic cells via DC-SIGN. Hum Reprod 30:1545-56. doi: 10.1093/humrep/dev113

83. Stax MJ, van Montfort T, Sprenger RR, Melchers M, Sanders RW, van Leeuwen E, Repping S, Pollakis G, Speijer D and Paxton WA (2009) Mucin 6 in seminal plasma binds DC-SIGN and potently blocks dendritic cell mediated transfer of HIV-1 to CD4(+) T-lymphocytes. Virology 391:203-11. doi: 10.1016/j.virol.2009.06.011

84. Camasses A, Bogdanova A, Shevchenko A and Zachariae W (2003) The CCT chaperonin promotes activation of the anaphase-promoting complex through the generation of functional Cdc20. Mol Cell 12:87-100.

85. Soues S, Kann ML, Fouquet JP and Melki R (2003) The cytosolic chaperonin CCT associates to cytoplasmic microtubular structures during mammalian spermiogenesis and to heterochromatin in germline and somatic cells. Exp Cell Res 288:363-73.

86. Redgrove KA, Anderson AL, Dun MD, McLaughlin EA, O’Bryan MK, Aitken RJ and Nixon B (2011) Involvement of multimeric protein complexes in mediating the capacitation-dependent binding of human spermatozoa to homologous zonae pellucidae. Dev Biol 356:460-74. doi: 10.1016/j.ydbio.2011.05.674

87. Jeon YJ, Kim DH, Jung H, Chung SJ, Chi SW, Cho S, Lee SC, Park BC, Park SG and Bae KH (2010) Annexin A4 interacts with the NF-kappaB p50 subunit and modulates NF-kappaB transcriptional activity in a Ca2+-dependent manner. Cell Mol Life Sci 67:2271-81. doi: 10.1007/s00018-010-0331-9

88. Iwasa T, Takahashi R, Nagata K and Kobayashi Y (2012) Suppression of MIP-2 or IL-8 production by annexins A1 and A4 during coculturing of macrophages with late apoptotic human peripheral blood neutrophils. Biochim Biophys Acta 1822:204-11. doi: 10.1016/j.bbadis.2011.10.013

89. Roy N and Majumder GC (1989) Purification and characterization of an anti-sticking factor from goat epididymal plasma that inhibits sperm--glass and sperm--sperm adhesions. Biochim Biophys Acta 991:114-22.

90. Frenette G, Legare C, Saez F and Sullivan R (2005) Macrophage migration inhibitory factor in the human epididymis and semen. Mol Hum Reprod 11:575-82. doi: 10.1093/molehr/gah197

91. Girouard J, Frenette G and Sullivan R (2009) Compartmentalization of proteins in epididymosomes coordinates the association of epidymyal proteins with the different functional structures of bovine spermatozoa. Biol Reprod 80:965-72. doi: 10.1095/biolreprod.108.073551

92. Gupta P, Leroux C, Patterson BK, Kingsley L, Rinaldo C, Ding M, Chen Y, Kulka K, Buchanan W, McKeon B and Montelaro R (2000) Human immunodeficiency virus type 1 shedding pattern in semen correlates with the compartmentalization of viral Quasi species between blood and semen. J Infect Dis 182:79-87. doi: 10.1086/315644

93. Anyanwu SI, Doherty A, Powell MD, Obialo C, Huang MB, Quarshie A, Mitchell C, Buchanan W and Newman GW (2018) Detection of HIV-1 and Human Proteins in Urinary Extracellular Vesicles from HIV+ Patients. Advances in Virology 2018:16. doi: 10.1155/2018/7863412

94. Osborne BJ, Sheth PM, Yi T, Kovacs C, Benko E, Porte C, Huibner S, Le AQ, Danroth R, Baraki B, Mazzulli T, Brumme ZL and Kaul R (2013) Impact of antiretroviral therapy duration and intensification on isolated shedding of HIV-1 RNA in semen. J Infect Dis 207:1226-34. doi: 10.1093/infdis/jit026
Table 1. Functional enrichments in CLU protein-protein interaction network as determined by STRING analysis (Top 5)

| Biological Process (GO)                                      | count in gene set | FDR     |
|--------------------------------------------------------------|-------------------|---------|
| regulation of proteolysis                                   | 9                 | 1.25e-08|
| humoral immune response                                     | 7                 | 1.46E-08|
| platelet degranulation                                      | 6                 | 2.04E-08|
| complement activation, classical pathway                    | 4                 | 1.40E-06|
| regulation of protein activation cascade                    | 4                 | 3.68e-06|

| Molecular Function (GO)                                     | count in gene set | FDR     |
|--------------------------------------------------------------|-------------------|---------|
| receptor activator activity                                 | 2                 | 0.0029  |
| growth factor receptor binding                               | 3                 | 0.0033  |
| low-density lipoprotein particle receptor binding            | 2                 | 0.0035  |
| heparin binding                                              | 3                 | 0.0035  |
| chemoattractant activity                                    | 2                 | 0.0037  |

| Cellular Component (GO)                                     | count in gene set | FDR     |
|--------------------------------------------------------------|-------------------|---------|
| platelet alpha granule lumen                                 | 6                 | 2.12e-10|
| secretory granule lumen                                      | 7                 | 7.39e-09|
| extracellular region                                         | 11                | 7.59e-09|
| membrane attack complex                                      | 3                 | 4.42e-07|
| extracellular space                                          | 7                 | 1.07e-05|

| KEGG Pathways                                                | count in gene set | FDR     |
|--------------------------------------------------------------|-------------------|---------|
| Complement and coagulation cascades                           | 4                 | 4.34e-06|
| Systemic lupus erythematosus                                 | 4                 | 4.48e-06|
| Prion diseases                                               | 3                 | 1.46e-05|
| Amoebiasis                                                   | 3                 | 0.00022 |
| Focal adhesion                                               | 3                 | 0.0015  |
Table 2. Top-10 KEGG pathways of 102 significantly enriched proteins in blood plasma as compared to seminal plasma of HIV- donors

| KEGG pathway                               | Number of genes | P value     |
|--------------------------------------------|-----------------|-------------|
| 1  Complement and coagulation cascades      | 29              | 0           |
| 2  Prion diseases                           | 9               | 4.13E-13    |
| 3  Staphylococcus aureus infection          | 10              | 9.59E-13    |
| 4  Systemic lupus erythematosus             | 11              | 4.19E-10    |
| 5  Pertussis                                | 8               | 1.76E-08    |
| 6  African trypanosomiasis                  | 3               | 0.00132944  |
| 7  Vitamin digestion and absorption        | 2               | 0.009818541 |
| 8  Amoebiasis                              | 3               | 0.024735829 |
| 9  Fat digestion and absorption             | 2               | 0.02726622  |
| 10 Chagas disease (American trypanosomiasis)| 3               | 0.027378825 |

Table 3. Top-10 KEGG pathways of 218 significantly enriched proteins in seminal plasma as compared to blood plasma of HIV- donors

| KEGG pathway                               | Number of genes | P value     |
|--------------------------------------------|-----------------|-------------|
| 1  Lysosome                                 | 20              | 6.79E-14    |
| 2  Glycolysis / Gluconeogenesis             | 12              | 2.87E-09    |
| 3  Antigen processing and presentation      | 11              | 1.56E-07    |
| 4  Other glycan degradation                 | 6               | 6.02E-07    |
| 5  Carbon metabolism                        | 12              | 1.25E-06    |
| 6  Protein processing in endoplasmic reticulum| 14          | 2.35E-06    |
| 7  Renin-angiotensin system                 | 6               | 3.03E-06    |
| 8  Glycosaminoglycan degradation            | 5               | 2.03E-05    |
| 9  Metabolic pathways                       | 43              | 3.29E-05    |
| 10 Biosynthesis of amino acids              | 8               | 7.30E-05    |
Table 4. Top-10 KEGG pathways of 91 significantly enriched proteins in blood plasma as compared to seminal plasma of HIV+ donors

| KEGG pathway                          | Number of genes | P value       |
|---------------------------------------|-----------------|---------------|
| Complement and coagulation cascades    | 27              | 0             |
| Prion diseases                        | 10              | 2.44E-15      |
| Staphylococcus aureus infection       | 10              | 4.41E-13      |
| Systemic lupus erythematosus          | 12              | 8.21E-12      |
| Pertussis                             | 8               | 9.72E-09      |
| African trypanosomiasis               | 4               | 4.71E-05      |
| Vitamin digestion and absorption      | 3               | 0.000349663   |
| Fat digestion and absorption          | 3               | 0.00171837    |
| Amoebiasis                            | 4               | 0.002676484   |
| Chagas disease (American trypanosomiasis) | 3         | 0.022751504   |

Table 5. Top-10 KEGG pathways of 246 significantly enriched proteins in seminal plasma as compared to blood plasma of HIV+ donors

| KEGG pathway                         | Number of genes | P value       |
|--------------------------------------|-----------------|---------------|
| Lysosome                             | 27              | 0             |
| Glycolysis / Gluconeogenesis         | 10              | 2.44E-15      |
| Carbon metabolism                    | 10              | 4.41E-13      |
| Metabolic pathways                   | 12              | 8.21E-12      |
| Antigen processing and presentation  | 8               | 9.72E-09      |
| Pyruvate metabolism                  | 4               | 4.71E-05      |
| Cysteine and methionine metabolism   | 3               | 0.000349663   |
| Galactose metabolism                 | 3               | 0.00171837    |
| Other glycan degradation             | 4               | 0.002676484   |
| Protein processing in endoplasmic reticulum | 3        | 0.022751504   |
### Table 6. GO and KEGG pathways of 28 DPPs in blood exosomes upon HIV-1 infection

| Biological Process (GO)                                    | count in gene set | FDR       |
|-----------------------------------------------------------|-------------------|-----------|
| 1  Protein activation cascade                             | 4                 | 1.07E-05  |
| 2  Humoral immune response                                | 4                 | 0.00063   |
| 3  Complement activation, lectin pathway                  | 2                 | 0.0015    |
| 4  Blood coagulation, intrinsic pathway                   | 2                 | 0.003     |
| 5  Coagulation                                            | 3                 | 0.0146    |
| 6  Killing of cells of other organism                     | 2                 | 0.0348    |
| 7  Response to stress                                     | 6                 | 0.0399    |
| 8  Antimicrobial humoral immune response mediated by      |                   |           |
|    Antimicrobial peptide                                  | 2                 | 0.0399    |
| 9  Platelet degranulation                                  | 2                 | 0.0441    |
| 10 Defense response                                       | 4                 | 0.0441    |

| Cellular Component (GO)                                    | count in gene set | FDR       |
|-----------------------------------------------------------|-------------------|-----------|
| 1  extracellular region                                    | 6                 | 0.0158    |
| 2  platelet alpha granule lumen                           | 2                 | 0.0158    |
| 3  extracellular region part                               | 5                 | 0.0158    |
| 4  extracellular space                                     | 4                 | 0.0208    |
| 5  extracellular region                                    | 6                 | 0.0158    |

| KEGG Pathways                                             | count in gene set | FDR       |
|-----------------------------------------------------------|-------------------|-----------|
| 1  Complement and coagulation cascades                     | 2                 | 0.0041    |
Table 7. GO and KEGG pathways of 12 DPPs in semen exosomes upon HIV-1 infection

| Biological Process (GO)                                      | count in gene set | FDR   |
|-------------------------------------------------------------|-------------------|-------|
| 1 Regulation of heterotypic cell-cell adhesion              | 3                 | 0.00015 |
| 2 Regulation of establishment of protein localization       | 6                 | 0.00015 |
| 3 Positive regulation of substrate adhesion-dependent cell spreading | 3                 | 0.00016 |
| 4 Regulation of protein transport                           | 5                 | 0.00092 |
| 5 Plasminogen activation                                    | 2                 | 0.0017  |
| 6 Positive regulation of heterotypic cell-cell adhesion     | 2                 | 0.0019  |
| 7 Platelet degranulation                                    | 3                 | 0.0024  |
| 8 Regulation of protein secretion                           | 4                 | 0.0026  |
| 9 Fibrinolysis                                              | 2                 | 0.003   |
| 10 Innate immune response-activating signal transduction    | 3                 | 0.0031  |

| Molecular Function (GO)                                     | count in gene set | FDR   |
|-------------------------------------------------------------|-------------------|-------|
| 1 Lipase inhibitor activity                                  | 2                 | 0.0033 |

| Cellular Component (GO)                                     | count in gene set | FDR   |
|-------------------------------------------------------------|-------------------|-------|
| 1 fibrinogen complex                                        | 2                 | 0.0012 |
| 2 endoplasmic reticulum lumen                               | 4                 | 0.0012 |
| 3 secretory granule lumen                                   | 4                 | 0.0012 |
| 4 cell surface                                              | 4                 | 0.0051 |
| 5 platelet alpha granule lumen                              | 2                 | 0.0087 |
| 6 extracellular space                                       | 4                 | 0.0169 |
| 7 endomembrane system                                       | 7                 | 0.0169 |
| 8 vesicle                                                   | 5                 | 0.0277 |
| 9 extracellular region                                      | 5                 | 0.0314 |
| 10 intracellular organelle part                              | 9                 | 0.0314 |
| KEGG Pathways                             | count in gene set | FDR  |
|-------------------------------------------|-------------------|------|
| Complement and coagulation cascades       | 2                 | 0.0094 |
| Platelet activation                       | 2                 | 0.0094 |
| IL-17 signaling pathway                   | 2                 | 0.0094 |
Figures and Figure legends:

Figure 1: Overview of the sample preparation, analysis, and biophysical characterization of body-fluid exosomes: a, Study pipeline and proteomic workflow from participant clinical group to sample collection, and data analysis. b, Hierarchical representation of the studied variables. Numbers to the right represent variable levels, including 1- clinical group (HIV- vs HIV+), 2- body-fluid type (blood vs semen), 3- protein compartmentalization (exosomes vs exosome-free fractions), and 4- donors (3 donors in each group). c, Physical characteristics of BE and SE, showing exosome concentration and size as determined by Nano Tracking Analysis. d, Quantitation of AChE activity of 50 μg BE and SE for each uninfected and HIV infected participant at time 0 and 30 min. Dotted line correspond to the AChE activity of PBS negative control which was unchanged at time 0 and 30 min. Data are reported as the individual mean of triplicate measurements and error bars are SD. Statistical significance for AChE activity was determined between time 0 min and 30 min for each donor by two-tailed paired Student’s t-test. **** = <0.0001, *** = <0.001, ** = <0.01, * = <0.05.
Figure 2: Comparison of proteins identified from healthy and HIV infected blood and seminal plasmas: a, 3-way Venn diagram showing the number of unique and common proteins between the SpC and AUC data sets identified in this study. b, 4-way Venn diagram showing the number of unique and common proteins identified in this study compared to the list of top-100 most abundant exosomal proteins in ExoCarta (October 16, 2019). c, and d, 4-way Venn diagrams showing the compartmentalization of autologous blood and semen proteomes from healthy and HIV-infected donors.
Figure 3: Sample clustering and data normalization: a, Double-clustering (Euclidean, average linkage) heatmap of SpC data showing efficient body-fluid (blood vs semen) based separation. b, Double-clustering (Euclidean, average linkage) heatmap depicting loosely clustered AUC data set without directionality. c, Linear regression showing correlation between SpC and AUC data sets for the mean of each of the 587 common proteins. Pearson R, R² and P-value are shown. d, Bland-Altman plot of the ratio SpC/AUC as a function of the average of the two methods showing the bias and agreement between the SpC and AUC data. The bias of 0.03637 unit is represented by the gap between the X axis, corresponding to zero differences, and the parallel line to the X axis at 0.03637 units. The red dotted lines are representations of the 95% limits of agreement, from -1.96 to +1.96. Hybrid data set showing integration of SpC and AUC data sets (SpC_Adjusted_AUC). Adjustments were made where peptides with detectable AUC, but undetectable spectral count were removed from the data set, resulting in a new list of 587 proteins. The analysis produced reliable visual representation of the data set revealing donor-, body-fluid-, and HIV- dependent differences.
**Figure 4: Distribution of exProteins common to autologous blood and semen:**
a, Four- and three-way Venn diagrams for identification of protein compartmentalization in exosomes and exosome-free plasma within each body-fluid for each donor, and for identification of common proteins to every compartment in healthy and HIV-infected groups. b, Two-way Venn analysis identifying proteins common to blood and semen after filtering out variabilities from the donor, clinical status, and body-fluid type. The common proteins are in large and bold font for the four- three- and two-way Venns. c, Heatmap showing enrichment pattern across four variables—body-fluid type, clinical group, compartmentalization, and donor for the 8 common proteins in b. d, Quantitation of the mean difference in enrichment score for the 8 common proteins. Differentially present proteins were determined using the 2-way ANOVA test by controlling the FDR (Benjamini and Hochberg). Adjusted p-values are shown in supplemental Table 2. Differences were considered to be significant when adjusted p-value < 0.05. e, Protein-protein interaction (PPI) network analyses of CLU using STRING database [1, 2] analyzed February 8, 2019. f, Western blot of CLU in BE and SE from HIV- and HIV+ samples.
Figure 5: Identification of exProteins enriched in blood and semen exosomes in the absence of HIV: 
a, Two-way Venn diagram of proteins in the exosome and exosome-free compartments showing donor-dependent variation in protein compartmentalization and identifying BE and SE enriched proteins. Exosome-specific proteins are in large and bold font. 
b, Three-way Venn analysis identifying proteins that are common to BE and SE in all donor samples. Common proteins for each exosome type is in large and bold font. 
c, Quantitative intensity profile of the SpC_adjusted_AUC identified differentially present proteins (DPP) in BE and SE. Statistics was performed by Two-way ANOVA with the Benjamini and Hochberg original FDR method. Adjusted p-values are shown in supplemental Table 3. Differences were considered to be significant when adjusted p-value < 0.05. 
d, Heatmap of BE and SE DDP revealing body-fluid specific differences with minor donor-dependent variabilities. e, Four-way Venn diagram integrating exosomal proteins from the SpC and the DPP data sets.
Figure 6: Profiling exProteins enriched in blood and semen exosomes from HIV infected participants: a, Two-way Venn diagram of proteins in the exosome and exosome-free compartments showing donor-dependent variation in protein compartmentalization and identifying BE and SE enriched proteins. Exosome-specific proteins are in large and bold font. b, Three-way Venn analysis identifying proteins from that are common to BE and SE in all donor samples. Common proteins for each exosome type is in large and bold font. c, Quantitative intensity profile of the SpC_adjusted_AUC identified differentially present proteins (DPP) in BE and SE. Statistics was performed by Two-way ANOVA with the Benjamini and Hochberg original FDR method. Adjusted p-values are shown in supplemental Table 4. Differences were considered to be significant when adjusted p-value < 0.05. d, Heatmap of BE and SE DDP revealing body-fluid specific differences with minor donor-dependent variabilities. e, Four-way Venn diagram integrating exosomal proteins from the SpC and the DPP data sets.
Figure 7: Proteomic Analysis of blood and semen exosome proteins altered by HIV infection: a-d, Two-way Venn diagrams of the HIV- and HIV+ BE and SE exclusive proteins identified in Figs. 5a, 6a. Common proteins for each exosome type are in large and bold fonts. e, f, Quantitative intensity profile and heatmap of DPP in BE. g, h, Quantitative intensity profile and heatmap of DPP in SE. Statistical analysis for e and g was performed with two-way ANOVA, with the original FDR method of Benjamini and Hochberg. Adjusted p-values are shown in supplemental Tables 5, 6. Differences were considered to be significant when adjusted p-value < 0.05. i) Heatmap of paralogous ANXAs and CCTs in BE and SE. j, Independent western blot validation of ANXA4 and CCT6A. k, Molecular Interaction Search Tool, MIST-based protein-protein interaction between ANXA4 and CCT6A showing possible interaction between ANXA4 and CCT6A through CDH1.
Figure 8: Predictive interaction between differentially present exosomal and HIV proteins: a, b, SE-ANXA4•HIV and SE-CCT6A•HIV protein-protein interaction was analyzed with the HIV•Human interaction database. c, Evaluation of HIV LTR promoter activation potential of TZM-bl cells treated with 200 μg BE and SE from HIV- (n=3) and HIV+ groups (n=3) for 24 hours. Data are reported as the group mean (HIV- or HIV+) from individuals measured in triplicate. Error bars are SD. Values are reported as relative light units (RLU). HIV-1 NL4.3 virus was used as positive control, and vehicle treated cells as negative control (dotted lines). d, Evaluation of HIV infection (RNA) in SUPT1 cells treated with 200 μg BE and SE from HIV- (n=3) and HIV+ (n=3) groups for 24 hours. Data are reported as the group mean (HIV- or HIV+) from individuals measured in duplicate. Values are reported as HIV RNA fold change in reference to uninfected cells. Error bars are SD. HIV-1 NL4.3 virus was used as positive control, and vehicle treated cells as negative control (dotted lines). Statistical significance for RLU and HIV RNA: Fold ∆ was determined between each group and background RLU or background signal, respectively, by two-tailed paired Student’s t-test. ***** = <0.0001, **** = <0.001, *** = <0.01, ** = <0.05. Fold change values in reference to HIV-uninfected cells set at 1 were used for statistical analysis of HIV RNA: Fold ∆.
Figure 9: Functional analysis of pathway-identified nucleic acid binding function of SE:

a, EMSA-based analysis of complex formation between BE and SE with NF-kB p65 sequence. 50 μg of exosomes were lysed with 0.1% Triton and incubated with Cy5-labelled NF-kB p65 probe in non-denaturing conditions. The mixture was separated on 5% non-denaturing PAGE. Lanes 1 and 2, are controls and correspond to probe alone and 200 μg BSA, respectively. Lanes 3-6 correspond to HIV-BE, HIV+BE, HIV-SE and HIV+SE, respectively. b, Additional controls for HIV-SE DNA binding shift. Lanes 1-3 correspond to Sp1, AP1 and IL1β Cy5-labelled probes. Lanes 4 and 5 correspond to NF-kB p65 probe in the presence of 50 and 100 times excess of cold probe, respectively. Lane 6 corresponds to a random Cy5-labelled primer. Gels are representative from at least three experiments which showed same trends.
Figure 10: HIV infection impairs the adhesive function of SE: a, b, Analysis of the cell clustering and adhesive abilities of BE and SE. U937 monocytic and Jurkat T cells were plated in exosome-treated tissue culture plate and wells were imaged 24 hours later prior to washing (a) and after 3X PBS washing (b) to assess the effect of BE and SE on cell clustering and adhesion, respectively. c, d, Quantitation of adhesion events in panel b. Significance was determined by one-way ANOVA test using vehicle as control (Sidak test) or by unpaired t-test (Welch test) for HIV- vs HIV+ comparisons. *, p<0.05; ****, p<0.0001; ns = not significant. Error bars are SD of 5 replicates. Experiment was repeated twice for U937 cells and three times for Jurkat cells with similar results. Scale bar = 200 μm.