

Increased levels of inflammatory cytokines in the female reproductive tract are associated with altered expression of proteases, mucosal barrier proteins, and an influx of HIV-susceptible target cells

Kelly B Arnold^{1,13}, Adam Burgener^{2,3,4,13}, Kenzie Birse^{2,3}, Laura Romas^{2,3}, Laura J Dunphy¹, Kamnoosh Shahabi⁵, Max Abou², Garrett R Westmacott⁶, Stuart McCorrister⁶, Jessie Kwatampora⁷, Billy Nyanga⁷, Joshua Kimani^{3,7}, Lindi Masson^{8,9}, Lenine J Liebenberg⁹, Salim S Abdool Karim^{9,10}, Jo-Ann S Passmore^{8,9,11}, Douglas A Lauffenburger¹, Rupert Kaul^{5,7,12,14} and Lyle R McKinnon^{5,7,9,14}

Elevated inflammatory cytokines (EMCs) at mucosal surfaces have been associated with HIV susceptibility, but the underlying mechanisms remain unclear. We characterized the soluble mucosal proteome associated with elevated cytokine expression in the female reproductive tract. A scoring system was devised based on the elevation (upper quartile) of at least three of seven inflammatory cytokines in cervicovaginal lavage. Using this score, HIV-uninfected Kenyan women were classified as either having EMC ($n = 28$) or not ($n = 68$). Of 455 proteins quantified in proteomic analyses, 53 were associated with EMC (5% false discovery rate threshold). EMCs were associated with proteases, cell motility, and actin cytoskeletal pathways, whereas protease inhibitor, epidermal cell differentiation, and cornified envelope pathways were decreased. Multivariate analysis identified an optimal signature of 16 proteins that distinguished the EMC group with 88% accuracy. Three proteins in this signature were neutrophil-associated proteases that correlated with many cytokines, especially GM-CSF (granulocyte-macrophage colony-stimulating factor), IL-1 β (interleukin-1 β), MIP-3 α (macrophage inflammatory protein-3 α), IL-17, and IL-8. Gene set enrichment analyses implicated activated immune cells; we verified experimentally that EMC women had an increased frequency of endocervical CD4⁺ T cells. These data reveal strong linkages between mucosal cytokines, barrier function, proteases, and immune cell movement, and propose these as potential mechanisms that increase risk of HIV acquisition.

INTRODUCTION

Although the biological process of inflammation is central to the initiation of an effective host immune response, HIV is able

to turn both inflammation and immune activation to its advantage.^{1,2} Immune activation was one of the earliest observed consequences of HIV infection,³ even before the

¹Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ²National HIV and Retrovirology Laboratory, JC Wilt Infectious Disease Research Centre, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada. ³Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada. ⁴Unit of Infectious Diseases, Department of Medicine Solna, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. ⁵Department of Medicine, University of Toronto, Toronto, Canada. ⁶National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada. ⁷Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya. ⁸Institute of Infectious Disease and Molecular Medicine (IDM), University of Cape Town, Cape Town, South Africa. ⁹Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa. ¹⁰Mailman School of Public Health, Department of Epidemiology, Columbia University, New York, USA. ¹¹National Health Laboratory Service, South Africa and ¹²University Health Network, University of Toronto, Toronto, Canada. Correspondence: LR McKinnon (Lyle.mckinnon@caprisa.org)

¹³These authors contributed equally to this work.

¹⁴These authors contributed equally to this work.

virus was fully characterized. HIV prefers to replicate in activated CD4⁺ T cells,⁴ and T-cell immune activation in some cohorts was a better predictor of mortality compared with plasma viral load.⁵ The early stages of HIV infection are characterized by inflammation and profound immune dysregulation in the gut mucosa,^{6,7} and genital inflammation at this stage also correlates with an increased plasma viral load.⁸ Even during treated HIV infection, residual inflammation markers are associated with increased AIDS mortality,⁹ and with persistent gut immune dysfunction and barrier disruption.¹⁰ Taken together, it is clear that inflammation is a key mediator of HIV pathogenesis.

The role of inflammation in HIV transmission has been more difficult to ascertain. Non-human primate studies suggest that mucosal Simian immunodeficiency virus exposure leads to an inflammatory cascade required for the establishment of productive viral infection.¹¹ Genital levels of the α -defensins, antimicrobial peptides that induce proinflammatory cytokine release by T cells,¹² were associated with increased HIV acquisition rates in both women¹³ and men.¹⁴ Genital coinfections have been associated with mucosal inflammation, but are not required for inflammation to be present, suggesting that inflammation owing to multiple causes enhances HIV risk. The levels of inflammatory cytokines and chemokines, which signal the presence of infection and recruit activated immune cells to the mucosa, are frequently used as biomarkers of inflammation in the female reproductive tract (FRT).^{15–18} As such, it might be expected that elevated mucosal cytokines would be correlated with increased rates of HIV acquisition. In the CAPRISA002 study, increased levels of proinflammatory cytokines were associated with increased rates of HIV acquisition.¹⁹ These results were subsequently validated in the CAPRISA004 tenofovir 1% gel trial, where a similar cytokine profile was a strong predictor of subsequent HIV acquisition.²⁰ In this analysis, a threefold increased risk of HIV infection was observed in women who had elevated levels of at least three proinflammatory cytokines (including macrophage inflammatory protein-1 α (MIP-1 α), interleukin-8 (IL-8), MIP-1 β , IL-1 β , IL-1 α , and tumor necrosis factor- α (TNF- α)) in cervicovaginal lavage (CVL).

Despite the strong association of inflammatory cytokines with HIV acquisition, the cellular and molecular mechanisms that mediate this risk remain largely unknown. Recent technological advances in mass spectrometry offer the promising opportunity to discover new mechanisms linking inflammatory cytokines and HIV risk in human cohorts, via the concomitant measurement of hundreds of proteins from small volume human mucosal samples.^{21–24} However, collection of data on this scale can lead to complexities of interpretation, as deciphering long “significant” protein lists or univariate analyses can miss important effects on pathways and perturbations in function categories. This can be mitigated in part by clustering and pathway analyses tools, which offer the opportunity to understand study group distinctions based on groups of biomarkers. Additionally, multivariate classification and regression techniques enable identification of minimal protein signatures associated with disease state based on

protein “patterns”, or covarying proteins.²⁵ These have been used previously by our group and others to gain insight into complex biological processes.²⁶

Here we analyzed the mucosal proteome associated with elevated levels of cytokines measured in the same specimens collected *ex vivo* from high-risk -uninfected women. This elevated mucosal cytokine pattern, which has been associated with both higher rates of HIV acquisition and is over-represented in women with bacterial inflammatory sexually transmitted infections,²⁷ was associated with profound differences in mucosal protein expression. Unsupervised hierarchical clustering of mucosal protein expression levels indicated that elevated cytokines were associated with visually distinct proteomes, and pathway analyses indicated perturbed protease activity, actin filament organization, and epidermal cell differentiation pathways. Multivariate data-driven analysis techniques found a unique signature of 16 proteins that best defined the elevated cytokine group. Three of 16 proteins were neutrophil-associated proteases, which in turn were significantly correlated to multiple inflammatory cytokines, and an increased frequency of cytobrush-derived cervical CD4⁺ T cells. Overall, these data provide a detailed characterization of mucosal cytokine pathways in the female genital tract, and propose several potential mechanisms of HIV susceptibility.

RESULTS

Participant characteristics

This study recruited participants from a dedicated female sex worker clinic in Nairobi, Kenya. We used a scoring system to classify participants ($n = 96$) having elevated mucosal cytokines on the basis of relative levels of proinflammatory FRT cytokines measured in CVL to identify extreme phenotypes. Women who had at least three CVL cytokines (TNF- α , IL-1 α , IL-8, MIP-3 α , RANTES (regulated on activation normal T cell expressed and secreted), IL-1 β , and MIP-1 β) in the upper quartile were defined as “elevated”; 29% (28/96) of study participants met this definition, and the rest were considered to be non-elevated controls. To verify the utility of this definition, we used unsupervised hierarchical clustering of CVL cytokine levels (**Figure 1**). This analysis confirmed that our elevated cytokine group tended to have increased levels of most of the cytokines in the multiplex panel, and a large proportion of these women clustered as having a distinct inflammatory cytokine phenotype.

We next compared selected demographic, behavioral, and reproductive variables between participants groups (**Table 1**). The only significant difference between groups was douching, with inflamed women less likely to douche (32% vs. 62%, $P = 0.013$). Because the substance used to douche may have important mucosal effects,²⁸ we further compared differences in douching practice between groups. “Soap and water” and “water only” were the most commonly reported. Although douching was less common in elevated inflammatory cytokine (EMC) vs. controls, EMCs were more likely to douche with soap (55% vs. 32%, $P = 0.25$), whereas EMCs were less likely

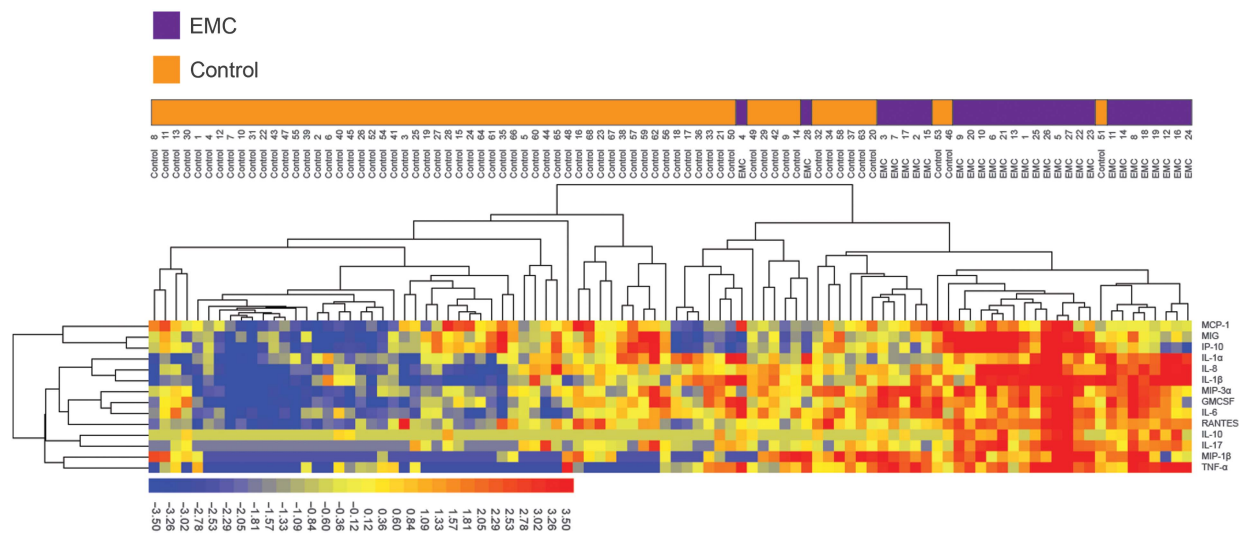


Figure 1 Unsupervised hierarchical clustering of individuals based on female reproductive tract (FRT) cytokines. Elevated inflammatory cytokine (EMC) (purple) and control (orange) groups were largely segregated using this approach, confirming that our cytokine scoring system identified women with a distinct cytokine phenotype.

Table 1 Participant characteristics

Variable	Elevated mucosal cytokines (% (n/N))	Controls (% (n/N))	P value
Age (median, IQR)	28.5 (25, 31)	28 (25.25, 34)	0.65
Marital status, divorced/widowed/separated ^a	60.7 (17/28)	54.4 (37/68)	0.65
Highest education = primary	82.1 (23/28)	82.4 (56/68)	0.35
Injectable contraceptives	21.4 (6/28)	35.3 (24/68)	0.23
Actively cycling ^a	85.2 (23/27)	73.1 (49/67)	0.29
Days since LMP if cycling (median, IQR)	10 (7, 20)	13 (9, 24)	0.25
Any STI (proportion)	14.3 (4/28)	7.4 (5/68)	0.44
HSV-2 serostatus	48.1 (13/27)	57.8 (37/64)	0.49
HPV infection	21.4 (6/28)	23.5 (16/68)	1.0
BV	20.8 (5/24)	6.9 (4/54)	0.11
Yeast on Gram stain	12.5 (3/24)	3.4 (2/58)	0.15
Number of clients per week (median, IQR)	8.5 (6, 11.5)	10 (7,14)	0.10
Condom use	71.4 (20/28)	55.9 (38/68)	0.18
Regular partner	32.1 (9/28)	16.2 (11/68)	0.10
Uses condoms with regular partner	78.3 (18/23)	47.1 (23/48)	0.13
Had sexual partners the day before the visit	50 (14/28)	52.9 (36/68)	0.97
Douche	32.1 (9/28)	61.8 (42/68)	0.013
Douche with water only ^b	44.4 (4/9)	53.7 (22/41)	0.721
Douche with water and soap ^c	55.6 (5/9)	31.7 (13/41)	0.253
Alcohol consumption	71.4 (20/28)	73.5 (50/68)	0.81

FRT, female reproductive tract; IQR, interquartile range; LMP, last menstrual period; STI, sexually transmitted infection.

^aRemainder were single.

^bDefined as <.35 days since LMP.

^cOf those who reported douching.

compared with controls to report douching with water only (44% vs. 53%, $P = 0.72$). It is important to note that this was not the objective of the study, and we were therefore underpowered

to draw any strong conclusions regarding the mucosal impact of vaginal douching. No differences in age, marital status, or education were observed. Women with bacterial vaginosis,

classical sexually transmitted infections, and/or vaginal yeast tended to be more inflamed, although the study was underpowered for this purpose. Similarly, trends were observed for having a regular sexual partner, the number of partners, and condom use, but none reached statistical significance ($P \geq 0.1$). The prevalence of herpes simplex virus and human papillomavirus infections at the study visit was comparable between groups. No differences in reproductive variables were evident.

Mucosal proteome associated with elevated female genital tract cytokines

We used proteomic analysis of CVL fluid to further examine differences between women. From a total of 455 human and microbial quantified factors in the FRT proteome, 109/455 (24.0%) were associated with elevated cytokines at the $P < 0.05$ level (Figure 2). Following multiple test correction based on a set false discovery rate (q -value) of 5% ($P < 0.0055$), 53 factors remained associated with elevated mucosal cytokines (Figure 2 and Supplementary Table 1 online). The top 9 differentially abundant factors (significant after Sidak–Bonferroni correction, $P < 0.000095$) were all human, and included overabundant factors filamin-A (FLNA; +2.3-fold change), neutrophil collagenase (MMP8; +2.6-fold change), matrix metalloproteinase-9 (MMP9; +3.1-fold change), coronin-1 A (COR1A; +2.6-fold change), fibrinogen β -chain (FGB; +2.37-fold change), plastin-2 (LCPI; +2.7 fold change); and underabundant factors antileukoproteinase (SLPI; –3.6-fold change), serine protease inhibitors Kazal-type 7 (SPINK7; –4.4-fold change), and Kazal-type 5 (SPINK5; –2.10-fold change). Intriguingly, two of the top proteins positively associated with elevated cytokines were neutrophil proteases

(MMP9 and MMP8), whereas three proteins most negatively associated with elevated cytokines were antiproteases (SPINK5, SPINK7, and SLPI). Other proteins associated with elevated cytokines were cytoskeletal and clotting elements (FLNA, COR1A, and FGB), suggestive of tissue degradation and/or remodeling pathways.

To better understand combinations of proteins associated with elevated mucosal cytokines, we separated the 53 significantly different protein factors into those that were increased (60%) or decreased (40%) in elevated cytokine participant group; and used hierarchical clustering within each of these groups to determine patterns that differentiated study groups. Differentially abundant proteins for both increased (Figure 3a) and decreased (Figure 3b) factors clearly discriminated participant groups. To further explore the functional implications of these differences, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID). This analysis found overabundant proteins in the elevated cytokine group to be significantly associated with protease activity, cell motility, cell activation, calcium ion-binding processes, and actin filament organization and binding (Figure 3a). Several of these are critical for leukocyte migration (such as ELANE, CORO1A, ITGB2, and MYH9). Conversely, decreased protein factors in the elevated cytokine group were significantly associated with antiprotease activity, epithelial differentiation, and keratinization, suggestive of impaired epithelial barrier integrity (Figure 3b). Interestingly, several of these factors were specific components of the cornified envelope, a layer of highly crosslinked insoluble proteins important for an effective physical barrier of the epidermis (SPR1A, SPR1B, SPRR3, INV, CSTA, SPINK5, FLG).²⁹

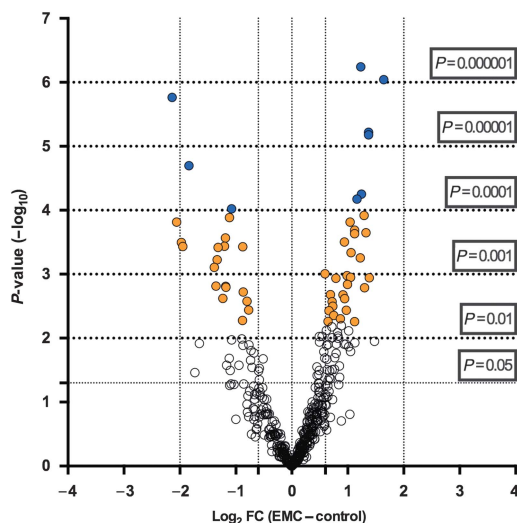


Figure 2 Volcano plot illustrating fold-change (FC) (x axis) and statistical significance distribution (y axis) of the proteomic data set. Data points in orange indicate proteins defined to be significantly differentially expressed based upon a set false discovery rate (q -value) of 5% ($P < 0.0055$). Proteins in blue denote those that reached statistical significance using the Sidak–Bonferroni method, with an α of 5% ($P < 0.000095$). These data indicate a wide distribution of both upregulated and downregulated proteins in mucosa in relation to elevated cytokines.

Multivariate model associates elevated cytokines with signatures of protease expression and cytoskeletal alterations

We next wanted to identify the minimum set of proteins that best differentiated participants with elevated cytokines from controls based on covariance, or relationships between proteins. To identify the minimum multivariate protein profile of elevated mucosal cytokines, we used the LASSO (Least Absolute Shrinkage and Selection Operator) method for regression and shrinkage and partial least-squares discriminant analysis (PLSDA).^{25,30} A profile of 16 proteins were best able to classify the women with elevated cytokines from controls. A PLSDA model of the 16 selected proteins provided excellent classification of study groups, with 88% calibration accuracy and 83% crossvalidation accuracy (Figure 4a). Latent variable 1 (LV1) was able to differentiate most participants with elevated cytokines (positive scores on LV1) from controls (negative scores on LV1; Figure 4b). Fifteen of the 16 identified markers were positively loaded on LV1, indicating that they were positively associated with elevated cytokines, whereas only one (cornulin) was negatively loaded on LV1, indicating that it was negatively associated with elevated cytokines. Interestingly, of these 16 biomarkers, 3 were proteases associated with neutrophils (MMP8, MMP9, and LYSC (lysozyme C)), whereas

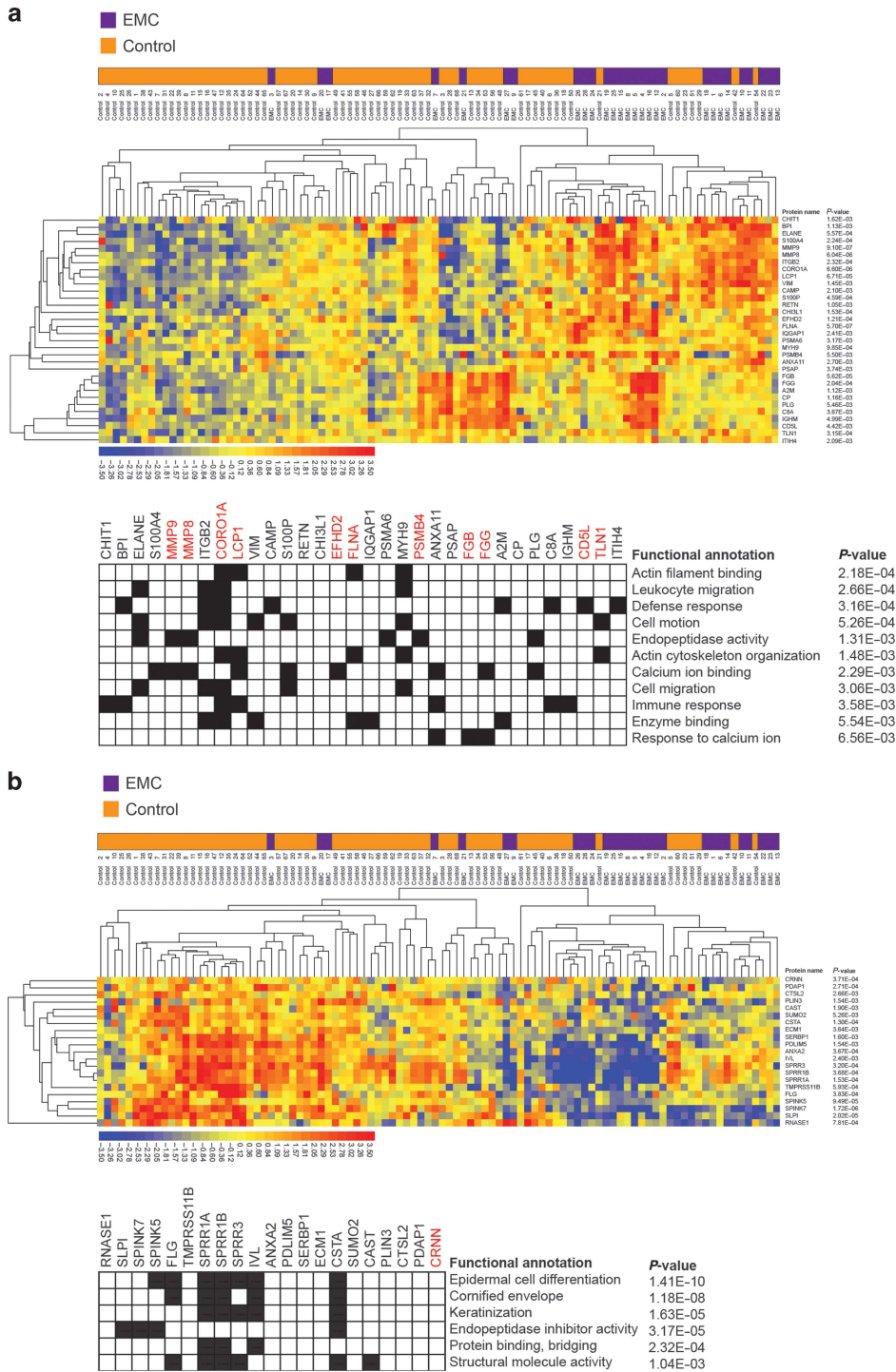


Figure 3 Heat map and functional analysis of differentially abundant proteins ($q < 0.05$) from cervicovaginal lavage samples between elevated inflammatory cytokine (EMC) and controls. Hierarchical clustering of proteins was generated by unsupervised average linkage using Pearson's correlation as the distance metric. The abundance of each protein is shown in color (red color denotes overabundant proteins, yellow unchanged, blue are underabundant compared to the mean) (color bar scale is beneath the figure). **(a)** Overabundant proteins in inflamed individuals and **(b)** underabundant proteins in EMC individuals. The highest scoring and significantly associated ($P < 0.05$) biological processes, molecular functions, and cellular components as identified by Database for Annotation, Visualization and Integrated Discovery (DAVID) gene ontology analysis with these proteins are shown below each heatmap. Not all associated functions are shown to reduce vertical sizing. Upregulated functions in the elevated cytokine group included mostly cell motility/activation and protease activity, whereas downregulated functions included antiprotease activity and epithelial differentiation/keratinization factors. LASSO (Least Absolute Shrinkage and Selection Operator) model proteins are indicated in red and their corresponding functional categories (denoted by asterisk), all belonging to the upregulated protein group as in **a**. Bar = fold increase, EMC/control.

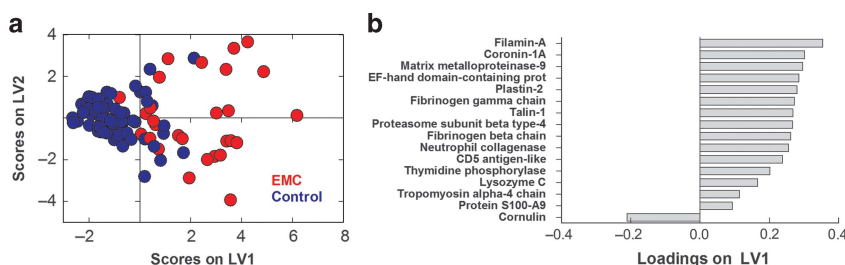


Figure 4 Protein signature associated with elevated mucosal cytokines. **(a)** LASSO (Least Absolute Shrinkage and Selection Operator) used covariance to identify a 16-protein signature that best classified elevated (red) and nonelevated (blue) groups, with 88% calibration accuracy and 83% crossvalidation accuracy. **(b)** Protein contributions to the identified signature can be visualized in the loadings plot, where positive loading indicates positive association with elevated cytokines, and negative loadings indicates comparative reduction in the elevated group. Three of 16 proteins in the identified signature were neutrophil proteases (matrix metalloproteinase 8 (MMP8), matrix metalloproteinase 9 (MMP9), and lysozyme (LYSC)), whereas others were associated with extracellular matrix (ECM), actin cytoskeleton, or cornified envelope. The LASSO identified signature was more predictive for study group status than each of 1000 different combinations of 16 other proteins for both calibration error ($P < 0.01$) and crossvalidation (CV) error ($P < 0.01$) (Supplementary Figure 1). LV1, latent variable 1.

most others were localized to the cell membrane, actin cytoskeleton, or extracellular matrix (CORO1A, FLNA, LCP1, TLN1, TPM4, FIBB, FIBG; Supplementary Table 1). As proteases and cytoskeletal proteins were all positively loaded on LV1, this suggests that these classes of proteins are together upregulated in conjunction with mucosal cytokine levels, potentially as part of the same biological process. We confirmed that our identified signature for classifying study groups was optimal by generating 1,000 additional PLSDA models, each with 16 different combinations of proteins selected from the remaining (non-LASSO) 439 measured proteins. This analysis suggested that our LASSO-selected signature was significantly better than random models for differentiating our study groups ($P < 0.01$), as both calibration accuracy and crossvalidation accuracy were 99.9th percentile rank compared with other models (Supplementary Figure 1). The fact that covariance of features were able to separate individuals based on having an elevated cytokine score suggests that relationships between the proteins in the identified signature may be of biological interest.

Elevated mucosal cytokine score was associated with gene sets from stimulated neutrophils and other resident immune cells in mucosa

To further understand potential cellular relationships with elevated mucosal cytokines, we used the gene set enrichment analysis toolset in an attempt to identify phenotypic associations with the inflamed proteomic data set.³¹ Our mucosal proteomic data was compared with independently generated data sets of various immune cell expression patterns. Interestingly, the top four gene sets included three positively associated and one negatively associated with having an elevated mucosal cytokine score. Interestingly, the top scoring gene set was that from stimulated neutrophils (11 proteins: normalized enrichment score = 2.10, $P < 0.0001$; Figure 5a). The next most enriched gene sets included stimulated CD8⁺ T cells (16 proteins: normalized enrichment score = 2.00, $P < 0.0001$; Figure 5b), pathogen-stimulated dendritic cells (11 proteins: normalized enrichment score = 1.96, $P < 0.0001$; Figure 5c), and influenza-stimulated dendritic cells (30

proteins: ES = 1.91, $P < 0.0001$; Figure 5d; full list in Supplementary Table 2). This influenza-stimulated DC signature could likely reflect converging downstream, secreted products that are result of Toll-like receptor or other pattern recognition of genitally relevant RNA viruses (but not necessarily influenza). The gene set that included the most LASSO identified factors was bacterial-stimulated peripheral blood mononuclear cells (MMP9, MMP8, LYSC, S100A9), consistent with findings suggesting bacterial infections are associated with elevated levels of many of the same cytokines in our panel.²⁷ These data suggest that increased activity of several immune cell subsets was associated with elevated FRT cytokines, and that this profile is consistent with increased microbial stimulation and/or responses.

An elevated mucosal cytokine score was associated with increased numbers of cervical CD4⁺ T cells

Although several levels of analysis suggested a relationship between elevated mucosal cytokines and increased frequency of immune cells, we wanted to validate this experimentally. While tissue data was not available at the time of this study, and neutrophils are technically challenging to stain in frozen samples, we were able to compare the absolute recovery of CD4⁺ T cells—the ideal target cell of HIV—in matched cervical cytobrush samples taken at the same study visit. Eleven (of 14) positive associations were observed between the cytokine level and the number of CD4⁺ T cells derived from cervical cytobrush sampling (Supplementary Table 3). Six of these associations were significant at $P < 0.01$ (Pearson's correlation), before adjusting for multiple test corrections. Similar results were obtained comparing the frequency of CD4⁺ T cells in women with cytokine concentrations in the upper quartile for a given cytokine, vs. the remainder of participants. Being in the upper quartile for MIP-3 α was the best predictor of cervical T cells ($P = 0.000317$). To test the relevance of our elevated mucosal cytokine score, we compared the frequencies of endocervical CD4⁺ T cells between study groups. Grouped as per our elevated cytokine score, EMC women had higher frequencies of CD4⁺ T cells compared with

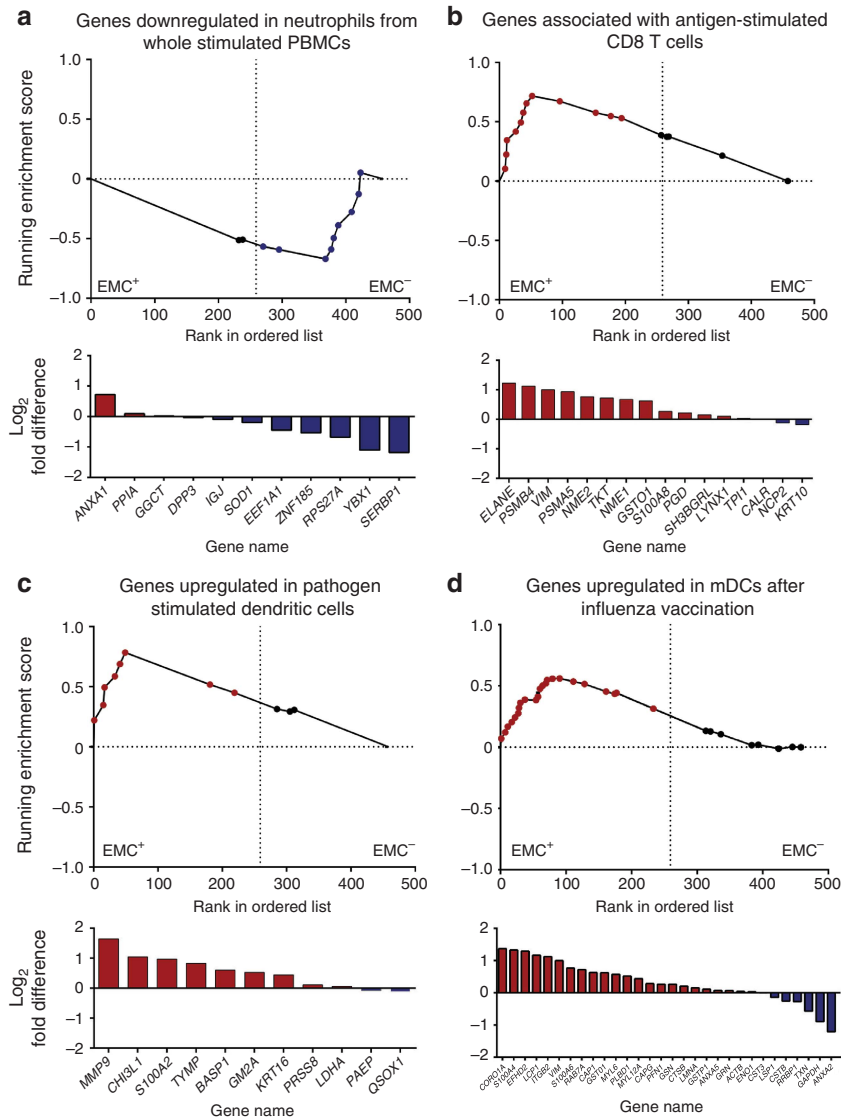


Figure 5 Immune cell gene sets associated with proteomic signatures of elevated mucosal cytokines. (a–d) This figure illustrates gene set enrichment plots of top stimulated immune cell types during elevated cytokines. Proteins contributing to the majority of the enrichment score are indicated by red dots (upregulated) in positively enriched gene sets in elevated cytokine individuals, and blue dots (downregulated) of negatively enriched gene sets in elevated cytokine individuals, and separated by a vertical dotted line. Corresponding gene sets are shown at the top of each figure. mDC, myeloid dendritic cell; PBMC, peripheral blood mononuclear cell.

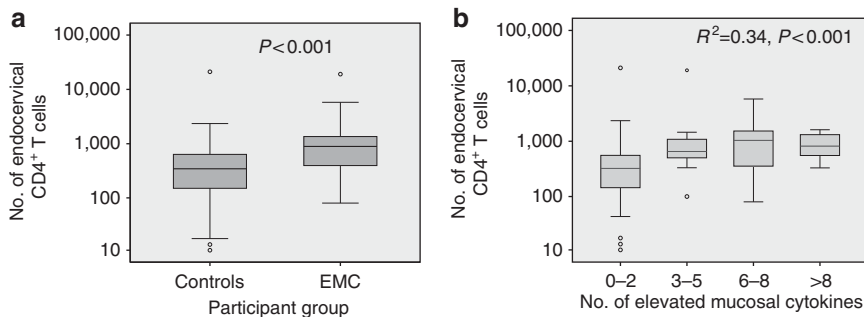


Figure 6 Significant increase in CD4⁺ T cells in endocervix of individuals with elevated mucosal cytokines. (a) Cytobrush measurements indicated a significant 2-fold increased in CD4⁺ T cells in female reproductive tract specimens of matched individuals in which three of seven inflammatory cytokines were elevated. *P* value determined using Mann–Whitney test. (b) Number of endocervical CD4⁺ T cells stratified by the number of elevated mucosal cytokines. ECM, extracellular matrix.

controls (median 896 vs. 345 cells, $P < 0.001$; **Figure 6a**). We also compared mucosal CD4⁺ T-cell frequencies with respect to increasing numbers of mucosal cytokines in the upper quartile; here, a marked increase in HIV target cells is evident in all participant groups with three or more elevated cytokines compared with women with two or fewer mucosal cytokines in the upper quartile (Pearson's $r^2 = 0.34$, $P < 0.001$; **Figure 6b**). These data confirm that elevated cytokines are associated with an increase in mucosal HIV target cells.

Protease levels are positively associated with cervical CD4⁺ T-cell numbers, and inflammatory cytokine expression

Although the unsupervised clustering suggests the colinearity of multiple cytokines elevated in the same samples, we wanted to know the individual cytokine contributions driving our main associations with neutrophil-associated proteases and epithelial ECM proteins. In particular, we sought to determine which of the inflammatory cytokines drive neutrophil activity, especially neutrophil proteases (MMP8, MMP9, and LYSC), which we hypothesize drove the signature of the elevated cytokine group. Expression of many inflammatory cytokines correlated positively with expression of MMP8, MMP9, and LYSC ($P < 0.01$ with adjustments for multiple comparisons; **Table 2**). Of all cytokines, IL-1 β , MIP-3 α , and IL-8 displayed the strongest correlation values, and correlated with all three proteases (all $P < 0.01$ after correction for multiple comparisons). There were also significant correlations between these three neutrophil-

associated proteins and cervical CD4⁺ T-cell numbers (**Table 2**).

DISCUSSION

Evidence from prospective human cohorts and nonhuman primate models suggest that inflammation may be central to the transmission of HIV and Simian immunodeficiency virus, but the precise mechanism(s) underpinning this association remain unclear. In this work, we used mass spectrometry measurements and various analytical methods to identify specific protein signatures most associated with inflammatory cytokines in cervicovaginal lavage samples. Interestingly, out of 455 proteins measured, we found that proteins involved in protease activity and epithelial barrier structures were the defining signature of women with elevated inflammatory cytokines. This was the main result from several different types of systems-level analysis, and presents a potential new mechanism inflammation-mediated HIV susceptibility.

In this work, the use of multiple system-level analytical techniques allowed us to establish that the genital proteomes of women with elevated cytokines were clearly distinguishable from controls. In standard univariate analysis, >10% of quantified proteins differed significantly between groups at the 5% false discovery rate level, and hierarchical clustering revealed visually distinct proteomes. Pathways analysis indicated several pathways were either up- (protease expression and cytoskeletal elements) or down- (antiprotease expression) regulated. LASSO and PLSDA allowed for identification of the minimum signature of proteins that differentiated EMC and control groups, and revealed a profile of 16 that differentiated the study groups with 88% accuracy, with more than 50% of the variance in the sample explained by this categorization. The advantage of these multivariate analysis techniques is that they enable identification of a key relationships between proteins that best separate study groups. Application of techniques such as this could provide significant insight into the mucosal immune environment that facilitates increased rates of HIV acquisition in prospective cohort studies of high-risk individuals. Further knowledge of the protein milieu that corresponds to elevated cytokines has broader relevance for understanding host defense in this immune compartment.

In identifying critical differences in the proteomes of women with EMC, our system-level analysis also generated new biological insight into processes most altered by inflammation. One key finding of our analysis was that that protein biomarkers of tissue remodeling and mucosal barrier integrity were strongly impacted by elevated cytokines, as was evident in all types of analyses that we used. Multiple proteins that regulate actin cytoskeleton organization (LCP1, TLN1, FLNA, CORIA, TPM4), several of which bind integrins and are concentrated at cell-cell contacts,³² were associated with elevated cytokines. Also present in the signature were ECM components (fibrinogen β - and γ -chain) that regulate cell adhesion and migration.³³ Previous work in mice indicated that inflammation induced ECM remodeling in dermal tissue, and this remodeling significantly altered immune cell trafficking.³⁴

Table 2 *R* values for Spearman's rank correlations of proteases, cytokines, and CD4⁺ T-cell numbers

	LYSC	MMP9	MMP8
GM-CSF	0.39*	0.50*	0.42*
IL-1 α	0.34	0.47*	0.46*
IL-8	0.42*	0.52*	0.52*
MCP-1	0.34	-0.13	-0.03
MIG	0.26	-0.01	0.11
MIP-3 α	0.45*	0.47*	0.50*
RANTES	0.34	0.29	0.47*
IL-10	0.24	0.37*	0.31
IL-17	0.37*	0.41*	0.45*
IL-1 β	0.45*	0.75*	0.67*
IL-6	0.30	0.59*	0.49*
IP-10	0.27	0.08	0.05
MIP-1 β	0.31	0.36*	0.31
TNF- α	0.15	0.60*	0.42*
No. of CD4 ⁺ cells	0.30*	0.48*	0.43*

GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin; IP-10, interferon gamma-induced protein 10; LYSC, lysozyme C; MIG, monokine induced by interferon- γ ; MIP, macrophage inflammatory protein; MMP, matrix metalloproteinase; RANTES, regulated on activation normal T cell expressed and secreted; TNF, tumor necrosis factor.

GM-CSF, IL-8, MIP-3 α , IL-17, and IL-1 β were all significantly correlated with 3/3 proteases after correction for multiple comparisons. All three proteases were also correlated with CD4⁺ T-cell number.

*Significance ($P < 0.01$) after correction for multiple comparisons.

However, specific cell types upstream of remodeling were not investigated. In contrast, the only downregulated protein in the PLSDA model was cornulin, a key marker of late epidermal differentiation and expressed in lower levels of cornified envelope.³⁵ Although not an integral part of the cornified envelope itself, this complements the univariate analysis showing other factors involved in the cornified envelope or in epidermal differentiation are downregulated. This layer is thought to be an important physical and biological barrier for host defense in the vaginal tract against microbial infections.³⁶

Taken together, these findings suggest a model in which tissue remodeling occurs at the expense of effective barrier function. The low per-coital rates of HIV transmission (1 in 1,250 exposures) suggest that the mucosal barrier in most cases is quite effective in preventing HIV infection, in the absence of transmission cofactors.^{37,38} Although inflammation and its associated cytokines, chemokines, and growth factors are required to fight infection and promote tissue repair, mucosal inflammation as a prolonged or inappropriate process could lead to impaired function of this effective barrier, as is observed in the HIV infected gut.^{39,40} This breakdown could provide HIV with the critical opportunity needed to gain access to the immune cells that are required to establish productive infection.

The second novel finding of our analysis tied to barrier function is the upregulation of several neutrophil-associated proteins. To date, no data to our knowledge have linked neutrophils and neutrophil-secreted proteins to the function of the mucosal barrier during HIV infection.⁴¹ Intriguingly, 3 of 16 proteins in the identified signature were neutrophil-associated proteases (lysozyme, neutrophil collagenase, and MMP9), whereas another (Protein S100-A9) has been linked to neutrophil chemotaxis and adhesion.⁴² In addition, neutrophil-associated gene set signatures were also identified within the proteomic data set, although these do not preclude the involvement of other immune cell types. Neutrophils have been widely reported to accumulate in the FRT during infections,^{43,44} but interestingly have also reported to have a critical role in endometrial degradation and repair during normal menstruation.^{45,46} Preliminary data suggest that neutrophils are the most common cell recovered from endocervical cytobrush sampling.⁴⁷ MIP-3 α , IL-1 β , and IL-8 in particular may be critical cytokines that influence neutrophil protease expression, as both were significantly correlated to protease levels. Previous work in cell culture indicates both IL-1 β and MIP-3 α induce neutrophil migration, and that IL-1 β is a defining feature of sexually transmitted infections.²⁷

Neutrophils have been associated with T helper type 17 cells,⁴⁸ which in turn are associated with both IL-1 β and MIP-3 α . T helper type 17 cells preferentially express CCR6, the receptor of MIP-3 α , and IL-1 β is key in driving the T helper type 17 phenotype.⁴⁹ Indeed, we also found that IL-17 was also significantly correlated with protease expression. Recently, our group showed that Th17 cells are depleted from the cervix of HIV-infected individuals,⁵⁰ and their number and function to be reduced in the gut.^{10,51} Another group has also recently shown these to be optimal HIV target cells.⁵² Furthermore,

MIP-3 α is increased in the cervix during early HIV infection,⁵³ and blockade of chemokines including MIP-3 α by anti-inflammatory glycerol monolaurate prevented nonhuman primates from acquiring Simian immunodeficiency virus infection.¹¹ However, it is worth noting that as we did not ascertain the cellular source of these protein factors, future experiments will be needed confirm that they originated from neutrophils vs. other immune cells of the mucosa. Unfortunately, the lavage specimens used here are typically insufficient for detailed cellular work.⁴⁷ Cytobrush samples from the endocervix may contain high concentrations of neutrophils,⁴⁷ however, because the matching cytobrush specimen for the current study was used to run the CD4⁺ T-cell panel, no cells were available for neutrophil quantitation. In addition, as neutrophils are highly prone to apoptosis, these would best measured in fresh samples collected as part of a new study. Despite this limitation, the evidence from the literature and data presented in our study warrant further evaluation of neutrophils in the context of HIV susceptibility presented in the paper, with a study specifically designed to overcome the technical challenges of study neutrophils *ex vivo*. It is also important to note that it could be neutrophil activities “per capita”, rather than overall neutrophil numbers, which underpin the associations we find with EMC.

One of the fundamental physiological roles of the cytokines often used to define inflammation is the recruitment of activated immune cells; in our study, several different system analyses of the mucosal proteome pointed to an association between elevated mucosal cytokines and increased immune cell frequency and activity. We then verified this by showing that women with elevated cytokines had significant increases in the number of endocervical CD4⁺ T cells. In one respect, this finding suggests that our inflammatory cytokine scoring method corresponds to inflammation as measured by the frequencies of immune cells. Although assessments of additional cell types and the location of cells within the tissue is a focus of ongoing work, the validity of quantifying CD4⁺ T cells has particular relevance to HIV acquisition risk, since these are now well established as the optimal viral targets at the time of transmission. These data also confirm previous work suggesting that proinflammatory cytokine levels were associated with an increase in cervical T cells.⁵⁴

One of the limitations of our study is the arbitrary use of a scoring system to determine participants with elevated cytokines. However, we believe this classification is useful for several reasons. For one, it places ~30% of participants into the “elevated” group, which we believe captures the “most inflamed” individuals; this is supported by unsupervised clustering as defining participants with upregulation of multiple cytokines, suggestive of a robust phenotype. Second, a similar scoring system was associated with HIV as an outcome in a prospective analysis of the CAPRISA004 study.²⁰ Although we cannot relate this mechanism to HIV susceptibility directly in this study, an expansion of this work has been planned. Third, our elevated cytokine classification captures a subclinical definition of inflammation that is in line with the common presentation of classical bacterial sexually transmitted infection

such as chlamydia and gonorrhea.¹⁹ Although inflammatory sexually transmitted infections only explain one aspect of stimuli that might lead to increased cytokine levels,⁵⁵ these serve as a model for defining a pathogenic mucosal immune state. Finally, our confirmatory cytobrush data found that three or more elevated cytokines appeared to be a threshold whereby the frequencies of HIV target cells markedly increased (Figure 6b). While the field of inflammation has grown to considerable complexity in recent years, our analysis using elevated cytokines to classify high-risk women sheds considerable light on the possible functions of these inflammatory cytokines in the female reproductive tract.

In summary, we found a set of mucosal proteins that defined distinct proteomes associated with elevated mucosal cytokines, with potential implications for what mechanisms might underpin the role these cytokines have in HIV risk. One hypothesis that emerges from our data links the covariance of neutrophil proteases with the ECM, actin cytoskeleton, and cornified envelope. In this model, inflammation-associated neutrophil proteases perturb epithelial cell differentiation, cell-cell contacts, and ultimately barrier function and integrity (Supplementary Figure 2). This could originate as a result of infection or another type of injury to tissue, causing resident cells such as epithelial, DC, and/or CD8⁺ T cells to release inflammatory cytokines.⁵⁶ This would then lead to immune cell influx, including neutrophils, which secrete proteases that would lead to further tissue injury, a decline in barrier function, and further inflammatory cytokine release. These results underscore that while elevated cytokines in the FRT are important for physiological processes such as menstruation, implantation and pathogen defense,^{8,57–60} these may be detrimental in the context of mucosal HIV exposure. A better understanding of the timing, causes, and consequences of vaginal inflammation might not only improve our understanding of female reproductive health but also help to prevent HIV acquisition in young women.

METHODS

Study population. All participants ($n = 96$) were HIV-uninfected female sex workers residing in Nairobi, Kenya accessing a clinic in the Kariobangi area. All participants gave informed written consent before participation and the study was approved at Institutional Review Boards at Kenyatta National Hospital (Kenya) and the University of Toronto (Canada). At the time of sampling, a questionnaire was administered capturing a range of demographic, reproductive, and behavioural variables. Participants were tested for a range of sexually transmitted infections including human papillomavirus (Aptima), herpes simplex virus-2 serostatus, HIV serostatus, *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, and bacterial vaginosis. All infections were treated as per Kenyan guidelines.

Flow cytometry. Endocervical cytobrushes were obtained as previously described by insertion and 360° rotation of a cytobrush in the endocervical os.⁵⁰ Cytobrushes were kept on ice until transport to the lab within 4 h. Cells were liberated from the brush through a combination of vortexing and washing, filtered, and cryopreserved for subsequent batched analysis. Thawed cells were washed two times and rested for 4 h, followed by staining to characterize CD4⁺ T-cell

populations. Panels of pretitrated antibodies and live-dead stain were incubated for 20 min at room temperature. Cells were then washed, fixed, and acquired on a BD LSR2 flow cytometer (Becton Dickinson, Franklin Lakes, NJ) configured for 10 colors. All events were collected, allowing an estimation of cells per cytobrush. For the purposes of this analysis, a gating strategy was used to define CD4⁺ T cells as those cells in the live, singlet, lymphocyte gate that coexpressed CD3 and CD4 (analyzed in FlowJo v.9.7.5).

Cytokine/chemokine measurements. CVLs were screened using the mesoscale discovery electrochemiluminescent ELISA system, and compared between participant groups. A panel of 14 markers was assessed (GM-CSF, IL-1 α , IL-8, MCP-1 (monocyte chemotactic protein-1), MIG (monokine induced by interferon- γ), MIP-3 α , RANTES, IL-10, IL-17, IL-1 β , IL-6, interferon γ -induced protein 10 (IP-10), MIP-1 β , TNF- α). CVL was plated at 50 μ l per well and run in duplicate. A standard curve was used to determine the concentration (pg ml^{-1}). The lower limit of quantitation was determined as the dilution in which the coefficient of variation exceeded 30%. Any sample above the range of the standard curve was repeated following dilution. Samples were run and analyzed by personnel who were blinded to study status. Data were presented as log₁₀-transformed values to normalize distributions.

Sample preparation for mass spectrometry analysis. Protein content of CVL samples was determined by BCA assay (Novagen, Etobicoke, ON, Canada). Equal protein amounts (100 μ g) from each sample were then individually denatured with urea exchange buffer (8 M urea (GE HealthCare, Mississauga, ON, Canada), 50 mM HEPES (Sigma, St Louis, MO), pH 8.0) for 20 min at room temperature placed into Nanosep filter cartridges (10 kDa). Samples were centrifuged, and reduced with 25 mM dithiothreitol (Sigma) for 20 min, then alkylated with 50 mM iodoacetamide (Sigma) for 20 min, followed by washes with 50 mM HEPES. Trypsin (Promega, Madison, WI) was added (2 μ g/100 μ g protein) and incubated at 37°C overnight in the cartridge. Peptides were eluted off the filter with 50 mM HEPES, and were dried via vacuum centrifugation. Reversed-phase liquid chromatography (high pH RP, Agilent 1200 series microflow pump, Water XBridge column) was used for desalting and detergent removal of peptides using a step-function gradient as described previously.²¹ Eluted fractions were dried via vacuum centrifugation and kept at -80°C until analyzed by mass spectrometry.

Mass spectrometry analysis. CVL peptide samples were analyzed by label-free tandem mass spectrometry as described previously.²¹ Briefly, peptide samples were injected into a nanoflow LC system (Easy nLC; Thermo Fisher, Waltham, MA) connected inline to a LTQ Orbitrap Velos (Thermo Fisher) mass spectrometer. Database searching was carried out with Mascot v2.4.0 (Matrix Science, Boston, MA) against UniProtKB/SwissProt (2013-02) Human- and Bacteria-only databases. Label-free protein abundance levels based on MS peak intensities were calculated using Progenesis LC-MS software (v4.0 Nonlinear Dynamics, Durham, NC). Complete details of liquid chromatography and mass spectrometry instrument settings are as described previously.²¹

Proteomic data analysis. Relative levels of protein abundance were calculated by dividing by average intensity across all samples, followed by log transformation (base 2). Statistical analysis between inflamed and non-inflamed groups was performed by Student's *t*-test (parametric). Only proteins that had an average covariance of <25% (455 proteins), as determined through measurements of standard reference sample run at 10 sample intervals (total 11 times), were used in downstream analysis to exclude proteins with higher technical measurement variability. Graphical representation of proteomic data was constructed using GraphPad Prism software (v6.0c, San Diego, CA). Differentially abundant proteins were those that passed a set false discovery rate threshold of 5% ($P < 0.0055$). Clustering of differentially

abundant proteins was generated by unsupervised average linkage hierarchical clustering using Pearson's correlation coefficient as the distance metric. The two major branches identified in the heatmap (overabundant/underabundant) were analyzed using the DAVID Bioinformatics Resource (6.7).^{61,62}

Biological function analysis. Biological/molecular functions and cellular component annotations were based on gene ontologies. A modified Fisher's exact *P*-value is calculated to determine the probability that the association between each protein in the data set and functional pathway is random. Functional categories with *P*-values < 0.05 and at least three proteins selected were considered to be associated with each branch in the cluster analysis.

Gene set enrichment analysis. We compared protein expression profiles to immune-related collections of gene sets from the Immunological Signatures database (C7.all.v4.0). A ranked list was generated from normalized protein abundance levels across all individuals. Gene set enrichment analysis (version 2.1.0) was run using default parameters. Gene sets with the top 10 most intense (+/-) normalized enrichment score were significant at $\alpha = 0.05$ (false discovery rate < 0.25 as per gene set enrichment analysis recommendation) and had > 10 proteins associated were said to be enriched in the inflammation phenotype.

Statistical analysis and modeling. We compared participant groups in Table 1 using χ^2 and Mann-Whitney *U*-test for categorical and continuous variables, respectively. Initial analysis of proteomics data was carried out by a combination of independent sample *T*-tests and fold change, with two different multiple test corrections (listed in the text). Correlations between cytokines and proteases were assessed using Spearman's correlation coefficients, with Bonferonni adjustments for multiple comparisons. In the validation cohort, we compared protease expression between groups using a one-tailed Mann-Whitney test.

Multivariate analysis. The minimum set of proteins necessary to distinguish study groups was determined with the LASSO method for regression shrinkage and selection, implemented using Matlab software (Mathworks, Natick, MA). *K*-fold crossvalidation determined the optimum value of the tuning parameter ("s"), such that the resulting model had the lowest possible mean squared error for prediction and associated features were chosen as the minimum set of biomarkers. PLSDA assessed prediction ability of LASSO-selected biomarkers for classifying participant groups. Data were normalized with mean centering and variance scaling, and crossvalidation was performed by iteratively excluding random subsets (in groups of 9–10 data points) during model calibration, and then using excluded data samples to test model predictions.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at <http://www.nature.com/mi>

ACKNOWLEDGMENTS

We thank the study participants and clinic staff at the Kenya AIDS Control Project. This work was supported by grants from the Canadian Institutes of Health Research (MOP-89983 to R.K.; OCB134115 to A.B. and L.M.); patient care activities and the purchase of antiretroviral drugs were funded by the President's Emergency Plan for AIDS Relief (PEPFAR). The Canadian Institutes of Health Research and the International Infectious Diseases and Global Health Training Program (to L.R.M.) and the Ontario HIV Treatment Network (to R.K.) provided salary support. G.W. is supported by the Public Health Agency of Canada. Ragon Institute Fellowship to K.B.A. and D.A.L. were partially supported by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the US Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

DISCLOSURE

The authors declared no conflict of interest.

© 2016 Society for Mucosal Immunology

REFERENCES

- Deeks, S.G., Tracy, R. & Douek, D.C. Systemic effects of inflammation on health during chronic HIV infection. *Immunity* **39**, 633–645 (2013).
- Douek, D.C., Roederer, M. & Koup, R.A. Emerging concepts in the immunopathogenesis of AIDS. *Annu. Rev. Med.* **60**, 471–484 (2009).
- Lane, H.C. *et al.* Abnormalities of B-cell activation and immunoregulation in patients with the acquired immunodeficiency syndrome. *N. Engl. J. Med.* **309**, 453–458 (1983).
- Zhang, Z.-Q. *et al.* Roles of substrate availability and infection of resting and activated CD4+ T cells in transmission and acute simian immunodeficiency virus infection. *Proc. Natl. Acad. Sci. USA* **101**, 5640–5645 (2004).
- Giorgi, J.V. *et al.* Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. *J. Infect. Dis.* **179**, 859–870 (1999).
- Brenchley, J.M. *et al.* CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J. Exp. Med.* **200**, 749–759 (2004).
- Mehandru, S. *et al.* Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *J. Exp. Med.* **200**, 761–770 (2004).
- Roberts, L. *et al.* Genital tract inflammation during early HIV-1 infection predicts higher plasma viral load set point in women. *J. Infect. Dis.* **205**, 194–203 (2012).
- Kuller, L.H. *et al.* Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. *PLoS Med.* **5**, e203 (2008).
- Kim, C.J. *et al.* Mucosal Th17 cell function is altered during HIV infection and is an independent predictor of systemic immune activation. *J. Immunol.* **191**, 2164–2173 (2013).
- Li, Q. *et al.* Glycerol monolaurate prevents mucosal SIV transmission. *Nature* **458**, 1034–1038 (2009).
- Vaschetto, R. *et al.* Role of human neutrophil peptides in the initial interaction between lung epithelial cells and CD4+ lymphocytes. *J. Leukoc. Biol.* **81**, 1022–1031 (2007).
- Levinson, P. *et al.* Levels of innate immune factors in genital fluids: association of alpha defensins and LL-37 with genital infections and increased HIV acquisition. *AIDS* **23**, 309–317 (2009).
- Hirbod, T. *et al.* HIV acquisition is associated with increased antimicrobial peptides and reduced HIV neutralizing IgA in the foreskin prepuce of uncircumcised men. *PLoS Pathogen.* **10**, e1004416 (2014).
- Fichorova, R.N. *et al.* Interleukin (IL)-1, IL-6, and IL-8 predict mucosal toxicity of vaginal microbicide contraceptives. *Biol. Reprod.* **71**, 761–769 (2004).
- Schwartz, J.L. *et al.* Fourteen-day safety and acceptability study of 6% cellulose sulfate gel: a randomized double-blind Phase I safety study. *Contraception* **74**, 133–140 (2006).
- Fichorova, R.N., Tucker, L.D. & Anderson, D.J. The molecular basis of nonoxynol-9-induced vaginal inflammation and its possible relevance to human immunodeficiency virus type 1 transmission. *J. Infect. Dis.* **184**, 418–428 (2001).
- Fichorova, R.N. Guiding the vaginal microbicide trials with biomarkers of inflammation. *J. Acquir. Immune Defic. Syndr.* **37**, S184–S193 (2004).
- Mlisana, K. *et al.* Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. *J. Infect. Dis.* **206**, 6–14 (2012).
- Masson, L. *et al.* Genital inflammation and the risk of HIV acquisition in women. *Clin. Infect. Dis.*; <http://dx.doi.org/10.1093/cid/civ298> (advance online publication, 21 April 2015).
- Burgener, A. *et al.* A systems biology examination of the human female genital tract shows compartmentalization of immune factor expression. *J. Virol.* **87**, 5141–5150 (2013).
- Burgener, A. *et al.* Comprehensive proteomic study identifies serpin and cystatin antiproteases as novel correlates of HIV-1 resistance in

- the cervicovaginal mucosa of female sex workers. *J. Proteome Res.* **10**, 5139–5149 (2011).
23. Burgener, A. *et al.* Salivary basic proline-rich proteins are elevated in HIV-exposed seronegative men who have sex with men. *AIDS* **26**, 1857–1867 (2012).
 24. Romas, L.M. *et al.* A comparative proteomic analysis of the soluble immune factor environment of rectal and oral mucosa. *PLoS One* **9**, e100820 (2014).
 25. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395 (1997).
 26. Benedict, K.F. & Lauffenburger, D.A. Insights into proteomic immune cell signaling and communication via data-driven modeling. *Curr. Top. Microbiol. Immunol.* **363**, 201–233 (2013).
 27. Masson, L. *et al.* Defining genital tract cytokine signatures of sexually transmitted infections and bacterial vaginosis in women at high risk of HIV infection: a cross-sectional study. *Sex. Transm. Infect.* **90**, 580–587 (2014).
 28. Low, N. *et al.* Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis. *PLoS Med.* **8**, e1000416 (2011).
 29. Dinh, M.H., Okocha, E.A., Koons, A., Veazey, R.S. & Hope, T.J. Expression of structural proteins in human female and male genital epithelia and implications for sexually transmitted infections. *Biol. Reprod.* **86**, 32 (2012).
 30. Lau, K.S. *et al.* *In vivo* systems analysis identifies spatial and temporal aspects of the modulation of TNF- α -induced apoptosis and proliferation by MAPKs. *Sci. Signal.* **4**, ra16 (2011).
 31. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
 32. Morley, S.C. The actin-bundling protein L-plastin: a critical regulator of immune cell function. *Int. J. Cell. Biol.* **2012**, 935173 (2012).
 33. Gonzalez, A.L., El-Bjeirami, W., West, J.L., McIntire, L.V. & Smith, C.W. Transendothelial migration enhances integrin-dependent human neutrophil chemokinesis. *J. Leukoc. Biol.* **81**, 686–695 (2007).
 34. Overstreet, M.G. *et al.* Inflammation-induced interstitial migration of effector CD4⁺ T cells is dependent on integrin α V. *Nat. Immunol.* **14**, 949–958 (2013).
 35. Contzler, R., Favre, B., Huber, M. & Hohl, D. Cornulin, a new member of the ‘fused gene’ family, is expressed during epidermal differentiation. *J. Invest. Dermatol.* **124**, 990–997 (2005).
 36. Anderson, D.J., Marathe, J. & Pudney, J. The structure of the human vaginal stratum corneum and its role in immune defense. *Am. J. Reprod. Immunol.* **71**, 618–623 (2014).
 37. Powers, K.A., Poole, C., Pettifor, A.E. & Cohen, M.S. Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. *Lancet Infect. Dis.* **8**, 553–563 (2008).
 38. Boily, M.-C. *et al.* Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect. Dis.* **9**, 118–129 (2009).
 39. Brenchley, J.M. *et al.* Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat. Med.* **12**, 1365–1371 (2006).
 40. Raffatellu, M. *et al.* Simian immunodeficiency virus-induced mucosal interleukin-17 deficiency promotes *Salmonella* dissemination from the gut. *Nat. Med.* **14**, 421–428 (2008).
 41. Casulli, S. & Elbim, C. Interactions between human immunodeficiency virus type 1 and polymorphonuclear neutrophils. *J. Innate Immun.* **6**, 13–20 (2014).
 42. Kolaczowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* **13**, 159–175 (2013).
 43. Carey, A.J. *et al.* Infection and cellular defense dynamics in a novel 17 β -estradiol murine model of chronic human group B streptococcus genital tract colonization reveal a role for hemolysin in persistence and neutrophil accumulation. *J. Immunol.* **192**, 1718–1731 (2014).
 44. Moore-Connors, J.M., Fraser, R., Halperin, S.A. & Wang, J. CD4⁺CD25⁺Foxp3⁺ regulatory T cells promote Th17 responses and genital tract inflammation upon intracellular *Chlamydia muridarum* infection. *J. Immunol.* **191**, 3430–3439 (2013).
 45. Kaitu’u-Lino, T.J., Morison, N.B. & Salamonsen, L.A. Neutrophil depletion retards endometrial repair in a mouse model. *Cell Tissue Res.* **328**, 197–206 (2007).
 46. Salamonsen, L.A., Zhang, J., Hampton, A. & Lathbury, L. Regulation of matrix metalloproteinases in human endometrium. *Hum. Reprod.* **15**, 112–119 (2000).
 47. McKinnon, L.R. *et al.* Optimizing viable leukocyte sampling from the female genital tract for clinical trials: an international multi-site study. *PLoS One* **9**, e85675 (2014).
 48. Weaver, C.T. & Hatton, R.D. Interplay between the TH17 and TReg cell lineages: a (co-)evolutionary perspective. *Nat. Rev. Immunol.* **9**, 883–889 (2009).
 49. Coccia, M. *et al.* IL-1 β mediates chronic intestinal inflammation by promoting the accumulation of IL-17A secreting innate lymphoid cells and CD4(+) Th17 cells. *J. Exp. Med.* **209**, 1595–1609 (2012).
 50. McKinnon, L.R. *et al.* Characterization of a human cervical CD4⁺ T cell subset coexpressing multiple markers of HIV susceptibility. *J. Immunol.* **187**, 6032–6042 (2011).
 51. Chege, D. *et al.* Sigmoid Th17 populations, the HIV latent reservoir, and microbial translocation in men on long-term antiretroviral therapy. *AIDS* **25**, 741–749 (2011).
 52. Rodriguez-Garcia, M., Barr, F.D., Crist, S.G., Fahey, J.V. & Wira, C.R. Phenotype and susceptibility to HIV infection of CD4⁺ Th17 cells in the human female reproductive tract. *Mucosal Immunol.* **7**, 1375–1385 (2014).
 53. McKinnon, L.R. *et al.* Early HIV-1 infection is associated with reduced frequencies of cervical Th17 cells. *J. Acquir. Immune Defic. Syndr.* **68**, 6–12 (2014).
 54. Kaul, R. *et al.* Genital levels of soluble immune factors with anti-HIV activity may correlate with increased HIV susceptibility. *AIDS* **22**, 2049–2051 (2008).
 55. Soumillon, M. *et al.* Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *BioRxiv*; <http://dx.doi.org/10.1101/003236> (advance online publication, 5 March 2014).
 56. Schenkel, J.M. *et al.* Resident memory CD8 T cells trigger protective innate and adaptive immune responses. *Science* **346**, 98–101 (2014).
 57. Tabibzadeh, S. & Sun, X.Z. Cytokine expression in human endometrium throughout the menstrual cycle. *Hum. Reprod.* **7**, 1214–1221 (1992).
 58. Henning, T.R. *et al.* Increased susceptibility to vaginal simian/human immunodeficiency virus transmission in pig-tailed macaques coinfecting with *Chlamydia trachomatis* and *Trichomonas vaginalis*. *J. Infect. Dis.* **210**, 1239–1247 (2014).
 59. Kelly, R.W., King, A.E. & Critchley, H.O. Cytokine control in human endometrium. *Reproduction* **121**, 3–19 (2001).
 60. Rametse, C.L. *et al.* Role of semen in altering the balance between inflammation and tolerance in the female genital tract: does it contribute to HIV risk? *Viral Immunol.* **27**, 200–206 (2014).
 61. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
 62. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).