Microbial function and genital inflammation in young South African women at high risk of HIV infection

Arghavan Alisoltani1,2, Monalisa T. Manhanzva1, Matthys Potgieter3,4, Christina Balle5, Liam Bell6, Elizabeth Ross6, Arash Iranzadeh3, Michelle du Plessis6, Nina Radzey1, Zac McDonald6, Bridget Calder4, Imane Allali3,7, Nicola Mulder3,8,9, Smritee Dabee1,10, Shaun Barnabas1, Hoyam Gamieldien1, Adam Godzik2, Jonathan M. Blackburn4,8, David L. Tabb8,11,12, Linda-Gail Bekker8,13, Heather B. Jaspan5,8,10, Jo-Ann S. Passmore1,8,14,15 and Lindi Masson1,8,14,16,17*

Abstract

Background: Female genital tract (FGT) inflammation is an important risk factor for HIV acquisition. The FGT microbiome is closely associated with inflammatory profile; however, the relative importance of microbial activities has not been established. Since proteins are key elements representing actual microbial functions, this study utilized metaproteomics to evaluate the relationship between FGT microbial function and inflammation in 113 young and adolescent South African women at high risk of HIV infection. Women were grouped as having low, medium, or high FGT inflammation by K-means clustering according to pro-inflammatory cytokine concentrations.

Results: A total of 3186 microbial and human proteins were identified in lateral vaginal wall swabs using liquid chromatography-tandem mass spectrometry, while 94 microbial taxa were included in the taxonomic analysis. Both metaproteomics and 16S rRNA gene sequencing analyses showed increased non-optimal bacteria and decreased lactobacilli in women with FGT inflammatory profiles. However, differences in the predicted relative abundance of most bacteria were observed between 16S rRNA gene sequencing and metaproteomics analyses. Bacterial protein functional annotations (gene ontology) predicted inflammatory cytokine profiles more accurately than bacterial relative abundance determined by 16S rRNA gene sequence analysis, as well as functional predictions based on 16S rRNA gene sequence data (p < 0.0001). The majority of microbial biological processes were underrepresented in women with high inflammation compared to those with low inflammation, including a Lactobacillus-associated signature of reduced cell wall organization and peptidoglycan biosynthesis. This signature remained associated with high FGT inflammation in a subset of 74 women 9 weeks later, was upheld after adjusting for Lactobacillus relative abundance, and was associated with in vitro inflammatory cytokine responses to Lactobacillus isolates from the same women. Reduced cell wall organization and peptidoglycan biosynthesis were also associated with high FGT inflammation.

(Continued on next page)
Background

Despite the large reduction in AIDS-related deaths as a result of the expansion of HIV antiretroviral programs, global HIV incidence has declined by only 16% since 2010 [1]. Of particular concern are the extremely high rates of HIV infection in young South African women [1]. The behavioral and biological factors underlying this increased risk are not fully understood; however, a key predisposing factor that has been identified in this population is genital inflammation [2–4].

We have previously shown that women with elevated concentrations of pro-inflammatory cytokines and chemokines in their genital tracts were at higher risk of becoming infected with HIV [2]. Furthermore, the protection offered by a topical antiretroviral tenofovir microbicide was compromised in women with evidence of genital inflammation [3].

The mechanisms by which female genital tract (FGT) inflammation increases HIV risk are likely multifactorial, and increased genital inflammatory cytokine concentrations may facilitate the establishment of a productive HIV infection by recruiting and activating HIV target cells, directly promoting HIV transcription and reducing epithelial barrier integrity [5–7]. Utilizing metaproteomics, Arnold et al. showed that proteins associated with tissue remodeling processes were upregulated in the FGT, while protein biomarkers of mucosal barrier integrity were downregulated in individuals with an inflammatory profile [7]. This suggests that inflammation increases tissue remodeling at the expense of mucosal barrier integrity, which may in turn increase HIV acquisition risk [7].

Conversely, women who have “optimal” vaginal microbiota, primarily consisting of Lactobacillus species, have low levels of genital inflammation and a reduced risk of acquiring HIV [13]. Lactobacilli and the lactic acid that they produce may actively suppress inflammatory responses and thus may play a significant role in modulating immune profiles in the FGT [14–16]. However, partly due to the complexity and diversity of the microbiome, the immunomodulatory mechanisms of specific vaginal bacterial species are not fully understood [17].

Further adding to this complexity is the fact that substantial differences in the cervicovaginal microbiota exist by geographical location and ethnicity and that the properties of different strains within particular microbial species are also highly variable [15, 18].

In addition to providing insight into host mucosal barrier function, metaproteomic studies of the FGT can evaluate microbial activities and functions that may influence inflammatory profiles [19]. A recent study highlighted the importance of vaginal microbial function by demonstrating that FGT bacteria, such as Gardnerella vaginalis, modulate the efficacy of the tenofovir microbicide by active metabolism of the drug [20]. Additionally, Zevin et al. found that bacterial functional profiles were associated with epithelial barrier integrity and wound healing in the FGT [21], suggesting that microbial function may also be closely linked to genital inflammation. The aim of the present study was to utilize both metaproteomics and 16S rRNA gene sequencing to improve our understanding of the relationship between microbial function and inflammatory profiles in the FGTs of South African women.

Results

This study included 113 young and adolescent HIV-uninfected women (aged 16–22 years) residing in Cape Town, South Africa [22]. Seventy-four of these women had samples and metadata available for analysis at a second time point 9 weeks later. Ten women had samples available at the second time point, but not the first, and were included to validate the functional signature identified in the primary analysis of this study. STI and BV prevalence were high in this cohort, with 48/113 (42%)
of the women having at least one STI and 56/111 (50%) of the women being BV positive by Nugent scoring (Additional file 1: Table S1). The use of injectable contraceptives, BV, and chlamydia were associated with increased genital inflammatory cytokines, as previously described [10, 23].

FGT metaproteome associates with BV and inflammatory cytokine profiles
Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was conducted on the total protein acquired from lateral vaginal wall swabs for all participants. Using principal component analysis (PCA), it was found that women with BV clustered separately from women who were BV negative according to the relative abundance of all host and microbial proteins identified (Additional file 1: Fig. S1a). Women who were considered to have high levels of genital inflammation based on pro-inflammatory cytokine concentrations (Additional file 1: Fig. S2) tended to cluster separately from women with low inflammation, while no clear grouping was observed by STI status or chemokine profiles (Additional file 1: Fig. S1b-d). After adjusting for potentially confounding variables including age, contraceptives, prostate-specific antigen (PSA), and co-infections, BV was associated with the largest number of differentially abundant proteins, followed by inflammatory cytokine profiles, while fewer associations were observed between chemokine profiles and protein relative abundance and none between STI status and protein relative abundance (Additional file 1: Fig. S1e). Individual STIs were also not associated with marked changes in the metaproteome, with Chlamydia trachomatis, Neisseria gonorrhoeae, Trichomonas vaginalis, and Mycoplasma genitalium associated with significant changes in zero, five, seven, and eleven proteins, respectively (after adjustment for potential confounders and multiple comparisons). However, this stratified analysis is limited by the small number of STI cases for some infections and the prevalence of co-infections.

Differences in taxonomic assignment using metaproteomics and 16S rRNA gene sequencing
Of the proteins identified in lateral vaginal wall swab samples, 38.8% were human, 55.8% were bacterial, 2.7% fungal, 0.4% archaeal, 0.09% of viral origin, and 2% grouped as “other” (not shown). However, as the majority of taxa identified had ≤2 proteins detected, a more stringent cutoff was applied to include only taxa with ≥3 detected proteins, or 2 proteins detected in multiple samples. The final curated dataset included 44% human, 55% bacterial (n = 81 taxa), and 1% fungal proteins (n = 13 taxa) (Fig. 1a–d; Additional file 2: Table S2). When the relative abundance of the most abundant bacterial taxa identified using metaproteomics was compared to the relative abundance of the most abundant bacteria identified using 16S rRNA gene sequencing, a large degree of similarity was found at the genus level. A total of 6/9 of the genera identified using 16S rRNA gene sequence analysis were also identified using metaproteomics (Fig. 1e, f; Additional file 1: Fig. S3). As expected, both approaches showed that multiple Lactobacillus species were less abundant in women with high versus low inflammation, while BV-associated bacteria (including G. vaginalis, Prevotella species, Megaspheara species, Sneathia amnii, and Atopobium vaginae) were more abundant in women with high inflammation compared to those with low inflammation (Additional file 2: Table S3). However, species-level annotation differed between the 16S rRNA gene sequencing and metaproteomics analyses, with metaproteomics identifying more lactobacilli at the species level, while Candidatus Lachnocurva vaginae (previously known as BVAB1) was not identified using metaproteomics. Additionally, the relative abundance of the taxa differed between 16S rRNA gene sequencing and metaproteomics analyses.

Protein profiles differ between women defined as having high, medium, or low FGT inflammation
A total of 449, 165, and 39 host and microbial proteins were differentially abundant between women with high versus low inflammation, medium versus high inflammation, and medium versus low inflammation, respectively (Additional file 2: Table S4). Microbial proteins that were more abundant in women with high or medium inflammation versus low inflammation were mostly assigned to non-optimal bacterial taxa (including G. vaginalis, Prevotella species, Megasphaera species, S. amnii, and A. vaginae). Less abundant microbial proteins in women with high inflammation were primarily of Lactobacillus origin (Additional file 1: Fig. S4). Of the human proteins that were differentially abundant according to inflammation status, 93 were less abundant and 132 were more abundant in women with high compared to those with low inflammation. Human biological process gene ontologies (GOs) that were significantly underrepresented in women with high versus low inflammation included positive regulation of apoptotic signaling, establishment of endothelial intestinal barrier, ectoderm development, and cornification [false discovery rate (FDR) adj. p < 0.0001 for all; Additional file 1: Fig. S5]. Significantly overrepresented pathways in women with high versus low inflammation included multiple inflammatory processes—such as chronic response to antigenic stimulus, positive regulation of IL-6 production, and inflammatory response (FDR adj. p < 0.0001 for all; Additional file 1: Fig. S5).
Fig. 1 (See legend on next page.)
Weighted correlation network analysis of microbial and host proteins identified five modules (clusters) representing co-correlations between microbial and host proteins (Fig. 2a, b). The yellow module included primarily L. iners proteins, while the turquoise module primarily consisted of L. crispatus and host proteins. The gray and brown modules consisted entirely of host proteins, and the blue module included non-optimal bacteria and host proteins. Pro-inflammatory cytokines correlated inversely with the Lactobacillus modules (yellow and turquoise) and positively with the gray, brown, and blue modules. Chemokines were significantly positively correlated with the brown and gray modules and inversely with the turquoise module (Fig. 2c). Blue, yellow, turquoise, and gray modules correlated with BV Nugent score, while BV showed no significant correlation with the brown module (Fig. 2d). The finding that the brown module did not include any microbial proteins and was not associated with BV or STIs suggests the presence of inflammatory processes independent of the microbiota. For STIs, no significant correlations (p value < 0.05) with any of the five modules were detected (Fig. 2d). To determine whether the co-correlations of microbial proteins with host proteins had functional meaning, we profiled the top biological processes of proteins identified using metaproteomics (Fig. 2a, b). The top functional pathways based on 16S rRNA gene sequence data, (iii) microbial protein relative abundance (determined using metaproteomics), and (iv) microbial molecular function (determined by aggregation of GOs of proteins identified using metaproteomics) for prediction of genital inflammation status (low, medium, and high groups) were evaluated using random forest analysis. The three most important species predicting genital inflammation grouping by 16S rRNA gene sequence data were Prevotella sanguinegens, A. vaginae, and L. iners (Fig. 3a). The top functional pathways based on 16S rRNA gene sequence data for the identification of women with inflammation were prolactin signaling, biofilm formation, and tyrosine metabolism (Fig. 3b). Similarly, the most important proteins predicting inflammation grouping by metaproteomics included two L. iners proteins (chaperone protein DnaK and pyrophosphate phospho-hydrolase), two A. vaginae proteins (glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase), Megasphaera genomosp. type 1 cold-shock DNA-binding domain protein, and L. crispatus ATP synthase subunit alpha (Fig. 3c). The top molecular functions associated with genital inflammation included oxidoreductase activity (of multiple proteins expressed primarily by lactobacilli, but also Prevotella
Fig. 2 (See legend on next page.)
and *Atopobium* species), beta-phosphoglucomutase activity (expressed by lactobacilli), and GTPase activity (of multiple proteins produced by mainly *Prevotella*, *G. vaginalis*, and some lactobacilli) as illustrated in Fig. 3d. These findings suggest that the functions of key BV-associated bacteria and lactobacilli are critical for determining the level of genital inflammation. It was also found that the out-of-bag (OOB) error rate distribution was significantly lower for molecular function, followed by protein relative abundance, functional prediction based on 16S rRNA gene sequence data, and then taxa relative abundance (Fig. 3e). This suggests that the activities of the bacteria present in the FGT play a critical role in driving genital inflammation, in addition to the presence and relative abundance of particular taxa.

**Differences in microbial stress response, lactate dehydrogenase, and metabolic pathways between women with high versus low FGT inflammation**

Microbial functional profiles were further investigated by comparing molecular functions, biological processes, and cellular component GO enrichment for detected proteins between women with low versus high inflammation using differential expression analysis, adjusting for potentially confounding variables including age, contraceptives, PSA, and STIs (Fig. 4; Additional file 2: Table S5). The majority of microbial molecular functions (90/107), biological processes (57/83), and cellular components (14/23) were underrepresented in women with high compared to low genital inflammation, with most of the differentially abundant metabolic processes, including peptide, nucleoside, glutamine, and glycerophospholipid metabolic processes, decreased in women with inflammation. The phosphoenolpyruvate-dependent sugar phosphotransferase system (a major microbial carbohydrate uptake system including proteins assigned to *Lactobacillus* species, *S. amnii*, *A. vaginae*, and *G. vaginalis*) was also underrepresented in women with evidence of genital inflammation compared to women with low inflammation (FDR adj. *p* < 0.0001; Fig. 4a). Additionally, 1-lactate dehydrogenase activity was inversely associated with inflammation (Additional file 2: Table S5) and both large and small ribosomal subunit cellular components were underrepresented in those with high FGT inflammation (Fig. 4b). The only metabolic processes that were overabundant in women with high versus low inflammation were malate and ATP metabolic processes. Overrepresented microbial biological processes in women with genital inflammation also included response to oxidative stress and cellular components associated with cell division (e.g., mitotic spindle midzone) and ubiquitination (e.g., Doa10p ubiquitin ligase complex, Hrd1p ubiquitin ligase ERAD-L complex, RQC complex; Fig. 4).

**Underrepresented Lactobacillus peptidoglycan, cell wall, and membrane pathways in women with high versus low FGT inflammation**

Among the top biological processes inversely associated with inflammation were cell wall organization, regulation of cell shape, and peptidoglycan biosynthetic process (all adj. *p* < 0.0001; Fig. 4a). The GO terms, regulation of cell shape, and peptidoglycan biosynthetic process included identical proteins (e.g., D-alanylalanine synthetase, D-alanyl-D-alanine-adding enzyme and Bifunctional protein GlmU), while cell wall organization included the same proteins plus D-alanyl carrier protein. Although proteins involved in these processes were only detected for *Lactobacillus* species and may thus be biomarkers of the increased relative abundance of lactobacilli, each remained significantly associated with inflammation after adjusting for the relative abundance of *L. iners* and non-*L. iners* species (determined using 16S rRNA gene sequencing), as well as STI status, PSA detection, and hormonal contraceptive use (*beta coefficient = −1.68; 95% confidence interval (CI) = −2.84 to −0.52; *p* = 0.004 for cell wall organization; *beta coefficient = −1.71; 95% CI = −2.91 to −0.53; *p* = 0.005 for both regulation of cell shape and peptidoglycan biosynthetic process). This suggests that the relationship between these processes and inflammation is independent of the relative abundance of *Lactobacillus* species and that the cell wall and membrane properties of lactobacilli may play a role in modulating FGT inflammation. Furthermore, we found that multiple cell wall and membrane cellular components, including the S-layer, peptidoglycan-based cell wall, cell
Fig. 3 (See legend on next page.)

a

| Genus                          | Log2FoldChange | p-value |
|--------------------------------|----------------|---------|
| Prevotella                     | 0.5            | 0.51    |
| Mycoplasma hominis             | 0.51           |         |
| Fastidiospila                  | 0.52           |         |
| Prevotella(OTU336)             | 0.52           |         |
| Megaplasma                     | 0.53           |         |
| Dialister microaerophilus      | 0.54           |         |
| Peptostreptococcus anaerobius  | 0.54           |         |
| Dialister                     | 0.54           |         |
| Ureaplasma parvum              | 0.65           |         |
| Gardnerella(OTTU7)             | 0.67           |         |
| Gemella asaccharolytica        | 0.67           |         |
| Lactobacillus                  | 0.67           |         |
| Corynebacterium                | 0.75           |         |
| Lachnivaginum BVA81            | 1.18           |         |
| Aerococcus christensenii       | 1.71           |         |
| Stealthia amnii                | 1.89           |         |
| Gardinerella vaginis           | 2.23           |         |
| Lactobacillus iners            | 2.4            |         |
| Atopobium vaginae              | 2.64           |         |
| Stealthia sanguinegens         |                |         |

b

| Pathway                          | Log2FoldChange | p-value |
|----------------------------------|----------------|---------|
| Sphingolipid metabolism          | 0.39           |         |
| Biosynthesis of ansamycins       | 0.42           |         |
| Neomycin kanamycin and gentamicin biosynthesis | 0.43 |         |
| Primary immunodeficiency         | 0.44           |         |
| Prodigiosin biosynthesis         | 0.45           |         |
| Chloroalkane and chloroalkene degradation | 0.46 |         |
| Tuberculosis                     | 0.47           |         |
| PPAR signaling                   | 0.49           |         |
| Epithelial cell signaling        | 0.52           |         |
| Alamine aspartate and glutamate metabolism | 0.52 |         |
| Photosynthesis                   | 0.52           |         |
| GABAergic synapse                | 0.53           |         |
| Oxidative phosphorylation        | 0.53           |         |
| Sulfur relay system              | 0.63           |         |
| Secondary bile acid biosynthesis | 0.64           |         |
| Primary bile acid biosynthesis   | 0.66           |         |
| Carbon metabolism                | 0.67           |         |
| Tyrosine metabolism              | 1.03           |         |
| Biofilm formation                | 1.32           |         |
| Prolactin signaling              | 1.63           |         |

c

| Gene                              | Log2FoldChange | p-value |
|-----------------------------------|----------------|---------|
| 3OS ribosomal protein S13         | 0.44           | 0.46    |
| 60 kDa chaperonin                 | 0.46           |         |
| Formate C-acetyltransferase       | 0.47           |         |
| Transcription elongation factor GreA | 0.47         |         |
| Elongation factor Tu              | 0.48           |         |
| Threonine—tRNA ligase             | 0.49           |         |
| Fructose-1,6-bisphosphate aldolase, class II | 0.53        |         |
| Phosphoglycerate kinase           | 0.56           |         |
| 2,3-bisphosphoglycerate           | 0.57           |         |
| Glutamine synthetase              | 0.59           |         |
| Triosephosphate isomerase         | 0.75           |         |
| Elongation factor Tu(GBPAS6)      | 0.75           |         |
| Actin, cytoplasmic 2              | 0.9            | 0.91    |
| Phosphocarrier, HPr family        | 0.92           |         |
| ATP synthase subunit alpha        | 1.05           |         |
| Pyrophosphatase phospho—hydrolase | 1.22           |         |
| Cold—shock DNA—binding domain protein | 1.28         |         |
| Phosphoglycerate kinase(AO0E09DXN4) | 1.72          |         |
| Glyceraldehyde—3-phosphate dehydrogenase | 1.87        |         |
| Chaperone protein DnaK            | 2.05           |         |

d

| Function                          | Log2FoldChange | p-value |
|-----------------------------------|----------------|---------|
| ATP binding                       | 0.47           | 0.47    |
| D—alanine—poly(phosphoribitol) ligase activity | 0.47 |         |
| rRNA binding                      | 0.52           |         |
| aminoacyl—tRNA editing activity   | 0.53           |         |
| adenylate kinase activity         | 0.54           |         |
| SS rRNA binding                   | 0.62           |         |
| GTP binding                       | 0.63           |         |
| ribosome binding                  | 0.63           |         |
| phosphoglucomannose mutase activity | 0.64         |         |
| translation elongation factor activity | 0.64      |         |
| structural constituent of ribosome | 0.65           |         |
| phosphoglycerate kinase activity  | 0.68           |         |
| amidase activity                  | 0.7            |         |
| DNA binding                       | 0.77           |         |
| adenylosuccinate synthase activity | 0.84          |         |
| hydrolase activity                | 1.18           |         |
| glutamate—ammonia ligase activity | 1.22           |         |
| GTPase activity                   | 1.51           |         |
| beta—phosphoglucomutase activity  | 2.92           |         |
| oxidoreductase activity           | 3.27           |         |

e

Fig. 3 (See legend on next page.)
wall, and were similarly underrepresented in women with high versus low FGT inflammation (Fig. 4b).

Using metaproteomic data from an independent group of 10 women, we similarly found that peptidoglycan biosynthetic process, cell wall organization, and regulation of cell shape GOs were underrepresented in women with high inflammation (Fig. 5a).

To further investigate whether the associations between FGT inflammation and Lactobacillus biological processes and cellular components were independent of Lactobacillus relative abundance, we compared these GOs in BV-negative women with Lactobacillus-dominant communities who were grouped (median splitting of samples based on the first principal component of nine pro-inflammatory cytokine concentrations) as having low (n = 20) versus high (n = 20) inflammation (Fig. S6). Interestingly, it was found that, even though only non-significant trends towards decreased relative abundance of lactobacilli determined by 16S rRNA gene sequencing (p = 0.73) and metaproteomics (p = 0.14) were observed, several previously identified GOs, including peptidoglycan-based cell wall (p = 0.01), cell wall organization (p = 0.03), and S-layer (p = 0.01), remained associated with the level of inflammation (Fig. S6). This suggests an association between these biological processes and cellular components that is not fully explained by the relative abundance of the lactobacilli themselves.

**In vitro confirmation of the role of Lactobacillus cell wall and membrane pathways in regulating genital inflammation**

As previous studies have suggested that the cell wall, as well as the cell membrane, may influence the immunomodulatory properties of Lactobacillus species and that peptidoglycan may be directly immunosuppressive [14, 24], we further investigated the relationships between Lactobacillus cell wall and membrane properties and inflammation using 64 Lactobacillus species that were isolated from the same women [25]. From these lactobacilli, the 22 isolates that induced the lowest levels of pro-inflammatory cytokine production by vaginal epithelial cells and the 22 isolates that induced the greatest cytokine responses were selected and their protein profiles evaluated using LC-MS/MS. Cell wall organization, regulation of cell shape, and peptidoglycan biosynthetic process that were associated with genital inflammation in the FGT metaproteomic analysis of vaginal swab samples were also significantly overrepresented in isolates that induced low compared to high levels of inflammatory cytokines in vitro (Fig. 5b–d). Furthermore, plasma membrane, cell wall, and membrane cellular components were similarly overrepresented in isolates that induced low versus high levels of inflammation (Fig. 5e–i). The cell wall cellular component included multiple proteins involved in adhesion (adhesion exoprotein, LPXTG-motif cell wall anchor domain protein, collagen-binding protein, mannose-specific adhesin, putative mucin binding protein, sortase-anchored surface protein, and surface anchor protein). The relative abundance of this cell wall cellular component correlated positively with the level of adhesion of the isolates to vaginal epithelial cells in vitro (rho = 0.30, p = 0.0476), which in turn correlated inversely with in vitro inflammatory cytokine production by these cells in response to the isolates [25]. Importantly, the cell wall organization biological process included multiple proteins involved in peptidoglycan biosynthesis, which has the potential to dampen immune responses in a strain-dependent manner [14, 26].

**Changes in FGT metaproteomic profiles over time**

To evaluate variations in FGT metaproteomic profiles associated with inflammatory profiles over time, we compared proteins, taxa, and biological process and cellular component GOs between two time points 9 weeks apart (n = 74). The grouping according to inflammatory cytokine profile was consistent at both visits for 41/74 (55%) participants, including 7/12 (58%) women who received antibiotic treatment between visits (Fig. 6a). The findings based on the second visit also confirmed the role of inflammation as one of the drivers of variation in complex metaproteome data, and similar patterns of data distribution according to inflammation status were observed at both visits (Fig. 6b). When the proteins, taxa, and functions that were most closely associated with inflammation grouping (high vs. low) at the first visit were
Fig. 4 (See legend on next page.)
evaluated at the second visit, overall profiles remained similar between visits (Fig. 6c–e). Significantly lower relative abundance of Lactobacillus species and increased non-optimal bacteria were observed in women with inflammation at the second visit (Fig. 6d). Similarly, cell wall organization, regulation of cell shape, and peptidoglycan biosynthetic process were among the top biological processes associated with low versus high inflammation at both visits (for other top GO terms, see Fig. 6e). Among women assigned to the same inflammatory profile group at both visits, the relative abundance of cell wall-related GO terms, malate metabolic process, and response to oxidative stress were highly consistent (Additional file 1: Fig. S7a-c). However, decreased levels of inflammation between visits were associated with an increase in the relative abundance of cell wall GOs and decreased response to oxidative stress and malate metabolic process GOs (Additional file 1: Fig. S7d-f). On the other hand, increased levels of inflammation were associated with decreased relative abundance of cell wall GOs and increased response to oxidative stress and malate metabolic process GOs (Additional file 1: Fig. S7g-i).

**Discussion**

Understanding biological risk factors for HIV acquisition is critical for the development of effective HIV prevention strategies. FGT inflammation, defined by elevated inflammatory cytokine levels, has been identified as a key factor associated with HIV infection risk. However, the causes and underlying mechanisms are not fully understood. In this study, we demonstrate that, while the presence of particular microorganisms, including non-optimal bacteria and Lactobacillus species, is important for modulating FGT inflammation, their activities and functions are likely of greater importance. We showed that the molecular functions of bacterial proteins predicted FGT inflammation status more accurately than taxa relative abundance determined using 16S rRNA gene sequencing, as well as functional predictions based on 16S rRNA gene sequencing analysis. The majority of microbial biological processes were underrepresented in women with high compared to low inflammation (57/83), including metabolic pathways and a strong signature of reduced Lactobacillus-associated cell wall organization and peptidoglycan biosynthesis. This signature remained associated with FGT inflammation after adjusting for the relative abundance of lactobacilli, as well as other potential confounders (such as hormone contraceptive use, semen exposure, STIs), and was also associated with inflammatory cytokine induction in vaginal epithelial cells in response to Lactobacillus isolates in vitro. We observed a high level of consistency between FGT metaproteomic profiles of the same women at two time points 9 weeks apart, particularly among those with the same level of inflammation at both visits. However, cell wall organization and peptidoglycan biosynthesis increased between visits in women who moved from a higher to a lower inflammation grouping and decreased in women who moved from a lower to a higher inflammation grouping.

While we found a large degree of similarity between metaproteomics and 16S rRNA gene sequence taxonomic assignment determined for the same women [27], important differences in the predicted relative abundance of most bacteria were observed. Differences in the relative abundance of certain taxa observed using metaproteomics and 16S rRNA gene sequence data are not surprising as taxonomic assignment from metaproteomic approaches also indicates functional activity (measured by the amount of expressed protein), rather than simply the abundance of a particular bacterial taxon. Additionally, others have noted that 16S rRNA gene sequencing does not have the ability to distinguish between live bacteria and transient DNA [28]. We also observed differences that are likely due to database limitations. For example, Candidatus Lachnocurva vaginae (previously known as BVAB1) was not identified in the metaproteomics analysis. This is due to the fact that this species was not present in the UniProt or NCBI databases [29]. However, the Clostridium and Ruminococcus species which were identified are likely to be Candidatus Lachnocurva vaginae since these taxa fall into the same Clostridiales order. Additionally, Clostridium and Ruminococcus species were more abundant in
**Fig. 5** (See legend on next page.)

- **a**
  - ribonucleoside-diphosphate reductase complex
  - S-layer
  - small ribosomal subunit
  - large ribosomal subunit
  - peptidoglycan-based cell wall
  - peptidoglycan biosynthetic process
  - cell wall organization
  - transcription, DNA-templated
  - PEP-PTS
  - ATP metabolic process
  - response to oxidative stress
  - regulation of cell shape
  - malate metabolic process

- **b** $p=0.0172^*$
  - Cell wall organization

- **c** $p=0.0272$
  - Regulation of cell shape

- **d** $p=0.0327$
  - Peptidoglycan biosynthetic process

- **e** $p=0.0002^*$
  - Plasma membrane CC

- **f** $p=0.0092^*$
  - Cell wall CC

- **g** $p=0.0327$
  - Membrane CC

- **h** $p=0.1372$
  - S-layer CC

- **i** $p=0.1372$
  - Peptidoglycan-based cell wall CC

*Low (n=5)*
*High (n=5)*

Lactobacillus isolates inducing low inflammatory responses (n=22)
Lactobacillus isolates inducing high inflammatory responses (n=22)
women with high compared to low inflammation, as expected for *Candidatus* Lachnocurva vaginae [10]. Metaproteomic analysis was able to identify lactobacilli to species level, while sequencing of the V4 region of the 16S rRNA gene had more limited resolution for this genus, as expected.

The finding that fungal protein relative abundance was substantially lower than bacterial protein relative abundance is not surprising as it has been estimated that the total number of fungal cells is orders of magnitude lower than the number of bacterial cells in the human body [30]. Although *Candida* proteins were detected, the curated dataset (following removal of taxa that had only 1 protein detected or 2 proteins detected in only a single sample) did not include any *Candida* species. This may be due to the relative rarity of fungal cells in these samples, and it would be interesting to evaluate alternative methodology for sample processing to better resolve this population in the future [31]. A limitation of microbial functional analysis using metaproteomics is that, at present, there is generally sparse population of functional information in the databases for the majority of the microbiome. Thus, the findings of this study will be biased toward the microbes that are better annotated. Nonetheless, microbial function was closely associated with genital inflammatory profiles and it is possible that this relationship may be even stronger should better annotation exist. Another limitation of this study is that the in vitro *Lactobacillus* analysis was conducted using a transformed cell line system that is a greatly simplified model and cannot recapitulate the complexity of the FGT and the microbiome. It is however significant to note that similar proteomic signatures were noted both in vitro and in vivo, regardless of this limitation. To enable researchers to further study associations between microbial composition and function with BV, STIs, and FGT inflammation by conducting detailed analysis of specific proteins, taxa, and GO terms of interest, we have developed the “FGT-DB” online application. The application is available at http://fgtdb.org/ and allows users to repeat the analyses presented here, as well as obtain additional information by analyzing individual proteins, species, and pathways.

A significant functional signature associated with FGT inflammation grouping included *Lactobacillus*-associated cell wall, peptidoglycan, and cell membrane biological processes and cellular components. Although these functions were linked exclusively to *Lactobacillus* species, the associations with FGT inflammation were upheld after adjusting for *Lactobacillus* relative abundance determined by 16S rRNA gene sequencing and were also evident in an analysis including only women with *Lactobacillus*-dominant microbiota. This profile was confirmed when these GOs were compared between *Lactobacillus* isolates that induced high versus low levels of cytokine production in vitro. Previous studies have suggested that the peptidoglycan structure, the proteins present in the cell wall, as well as the cell membrane, may influence the immunomodulatory properties of *Lactobacillus* species [14, 24]. In a murine model, the administration of peptidoglycan extracted from gut lactobacilli was able to rescue the mice from colitis [26]. This activity was dependent on the NOD2 pathway and correlated with an upregulation of the indoleamine 2,3-dioxygenase immunosuppressive pathway. The immunomodulatory properties of peptidoglycan were furthermore dependent on the strain of lactobacilli from which the peptidoglycan was isolated [26]. The increase in cell wall organization and peptidoglycan biosynthesis that accompanied a reduction in local inflammation suggests an increase in *Lactobacillus* strains producing relatively large amounts of proteins involved in these processes. In a previous study, we showed that adhesion of lactobacilli to vaginal epithelial cells in vitro was inversely associated...
Fig. 6 (See legend on next page.)
with cytokine responses, suggesting that direct interaction between the isolates and vaginal epithelial cells is critical for immunoregulation [25]. In addition to cell wall and membrane properties, L-lactate dehydrogenase activity was identified as a molecular function inversely associated with inflammation. Similarly, D-lactate dehydrogenase relative abundance and D-lactate production by the Lactobacillus isolates included in this study were also associated with inflammatory profiles in an in vitro analysis [25]. These findings support the results of previous studies showing that lactic acid has immunoregulatory properties [16]. Taken together, these findings suggest that lactobacilli may modulate the inflammatory environment in the FGT through multiple mechanisms. It was further found that L. crispatus was more strongly associated with low inflammation in the differential abundance analysis and that L. iners was the Lactobacillus species most frequently detected in women with high inflammation. The co-expression analysis revealed that the majority of L. crispatus proteins grouped separately to L. iners proteins. However, both the L. crispatus and the L. iners modules were strongly inversely associated with inflammatory cytokine and chemokine concentrations. This is interesting since L. crispatus dominance is considered optimal, while the role of L. iners, the most prevalent Lactobacillus species in African women [10, 11], is poorly understood and it has been associated with compositional instability and transition to a non-optimal microbiota, as well as increased risk of STI acquisition [32, 33].

In this study, we observed a similar host proteome profile associated with high FGT inflammation compared to previous studies [7]. Multiple inflammatory pathways were overrepresented in women with high versus low FGT inflammation, while signatures of reduced barrier function were observed in women with high inflammation, with underabundant endothelial, ectoderm, and tight junction biological processes. These findings similarly suggest that genital inflammation may be associated with epithelial barrier function.

**Conclusions**

The link between FGT microbial function and local inflammatory responses described in this study suggests that both the presence of specific microbial taxa in the FGT and their properties and activities likely play a critical role in modulating inflammation. Currently, the annotation of microbial functions is sparse, but with ever-increasing amounts of high-quality, high-throughput data, the available information will improve steadily. The findings of the present study contribute to our understanding of the mechanisms by which the microbiota may influence local immunity, and in turn alter the risk of HIV infection. Additionally, the analyses described herein identify specific microbial properties that may be harnessed for biotherapeutic development.

**Methods**

**Participants**

This study included sexually experienced HIV-negative adolescent girls and young women (16–22 years) from the Women’s Initiative in Sexual Health (WISH) study in Cape Town, South Africa [22]. While the parent study cohort comprised 149 women, the present sub-study included 113 women who had both vaginal swabs available for metaproteomics analysis and menstrual cup (MC) cervicovaginal secretions available for cytokine profiling. Seventy-four of these women had specimens available at a second time point 9 weeks later (interquartile range 9–11 weeks) for metaproteomics analysis. Of these women, 12 received antibiotic treatment for laboratory-diagnosed or symptomatic STIs and/or BV (doxycycline, metronidazole, azithromycin, cefixime, ceftriaxone, and/or amoxicillin) between study visits, while 2 received topical antifungal treatment.
(clotrimazole vaginal cream). Ten additional women who did not have visit 1 samples available for analysis were included for validation purposes. The University of Cape Town Faculty of Health Sciences Human Research Ethics Committee (HREC REF: 267/2013) approved this study. Women ≥18 years provided written informed consent, and those <18 years provided informed assent and informed consent was obtained from parents/guardians.

**Definition of genital inflammation based on cytokine concentrations**

For cytokine measurement, MCs (Softcup®, Evaron Inc, USA) were inserted by the clinician and kept in place for an hour, after which they were transported to the lab at 4 °C and processed within 4 h of removal from the participants. MCs were centrifuged, and the cervicovaginal secretions were resuspended in phosphate-buffered saline (PBS) at a ratio of 1 ml of mucus: 4 ml of PBS and stored at −80 °C until cytokine measurement. Prior to cytokine measurement, MC secretions were pre-filtered using 0.2-μm cellulose acetate filters (Sigma-Aldrich, MO, USA). The concentrations of 48 cytokines were measured in MC samples using Luminex (Bio-Rad Laboratories Inc®, CA, USA) [23]. K-means clustering was used to identify women with low, medium, and high pro-inflammatory cytokine [interleukin (IL)-1α, IL-1β, IL-6, IL-12p40, IL-12p70, tumor necrosis factor (TNF)-α, TNF-β, TNF-related apoptosis-inducing ligand (TRAIL), interferon (IFN)-γ] and chemokine profiles [cutaneous T cell-attracting chemokine (CTACK), eotaxin, growth regulated oncogene (GRO)-α, IL-8, IL-16, IFN-γ-induced protein (IP)-10, monocyte chemoattractant protein (MCP)-1, MCP-3, monokine induced by IFN-γ (MIG), macrophage inflammatory protein (MIP)-1α, MIP-1β, regulated on activation, normal T cell expressed and secreted (RANTES)] (Additional file 1: Fig. S2).

**STI and BV diagnosis and evaluation of semen contamination**

Vulvovaginal swab samples were screened for common STIs, including *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, herpes simplex virus (HSV) types 1 and 2, *Treponema pallidum*, and *Haemophilus ducreyi*, using real-time multiplex PCR assays [22]. Lateral vaginal wall swabs were collected for BV assessment by Nugent scoring. Blood was collected for HIV rapid testing (Alere Determine™ HIV-1/2 Ag/Ab Combo, Alere, USA). PSA was measured in lateral vaginal wall swabs using Human Kalikrein 3/PSA Quantikine ELISA kits (R&D Systems, MN, USA).

**Discovery metaproteomics and analysis**

For shotgun LC-MS/MS, lateral vaginal wall swabs were collected, placed in 1 ml PBS, transported to the laboratory at 4 °C, and stored at −80 °C. Samples were computationally randomized for processing and analysis. The stored swabs were thawed overnight at 4 °C before each swab sample was vortexed for 30 s, and the mucus scraped off the swab on the side of the tube and vortexed for an additional 10 s. The samples were then clarified by centrifugation, and the protein content of each supernatant was determined using the Quanti-Pro bicinchoninic acid (BCA) assay kit (Sigma-Aldrich, MO, USA). Equal protein amounts (100 μg) were denatured with urea exchange buffer, filtered, reduced with dithiothreitol, alkylated with iodoacetamide, and washed with hydroxyethyl piperazineethanesulfonic acid (HEPES). Acetone precipitation/formic acid (FA; Sigma-Aldrich, MO, USA) solubilization was added as a further sample clean-up procedure. The samples were then incubated overnight with trypsin (Promega, WI, USA), and the peptides were eluted with HEPES and dried via vacuum centrifugation. Reversed-phase liquid chromatography was used for desalting using a step-function gradient. The eluted fractions were dried via vacuum centrifugation and kept at −80 °C until analysis. LC-MS/MS analysis was conducted on a Q-Exactive Quadrupole-Orbitrap MS (Thermo Fisher Scientific, MA, USA) coupled with a Dionex UltiMate 3000 nano-UPLC system (120 min per sample). The solvent system included solvent A: 0.1% formic acid (FA) in LC grade water (Burdick & Jackson, NJ, USA), and solvent B: 0.1% FA in acetonitrile (ACN; Burdick & Jackson, NJ, USA). The MS was operated in positive ion mode with a capillary temperature of 320 °C, and 1.95 kV electrospray voltage was applied. All 187 samples were evaluated in 11 batches over a period of 3 months. A quality control consisting of pooled lateral vaginal swab eluants from the study participants was run at least twice with every batch. Preliminary quality control analysis was conducted, and flagged samples/batches were rerun.

Due to the important role of database in metaproteomic analysis, we compared two different databases to identify proteins in our dataset: (i) UniProt database restricted to human and microbial entries (73,910,451, release August 2017) and filtered using the MetaNovo pipeline [34] and (ii) vaginal metagenome-based database with human proteins obtained from the study of Afuni-Zadeh et al. [35]. The raw data were processed using MaxQuant version 1.5.7.4 against each database, separately. The detailed parameters are provided in Additional file 1: Table S6. In brief, methionine oxidation and acetylation of the protein N-terminal amino acid were considered as variable modifications and carbamidomethyl (C) as a fixed modification. The digestion
enzyme was trypsin with a maximum of two missed cleavages. The taxonomic analysis of the proteins identified using both of the databases was comparable; however, the first database (filtered UniProt human and microbial entries) resulted in the identification of a greater number of proteins compared to the vaginal metagenome database, while exhibiting a high degree of similarity compared to the 16S rRNA gene sequence data. Therefore, the data generated using the first database was used for downstream analysis.

For quality control analysis, raw protein intensities and log10-transformed iBAQ intensities of the quality control pool, as well as the raw protein intensities of the clinical samples, were compared between batches (not shown). As variation in raw intensity was observed, all data were adjusted for batch number prior to downstream analysis. The cleaned and transformed iBAQ values were used to identify significantly differentially abundant proteins according to inflammatory cytokine profiles, BV status, or STIs using the limma R package [36]. The p values were obtained from the moderated t test after adjusting for confounding variables including age, contraceptives, PSA, and STIs. Proteins with FDR-adjusted p values < 0.05 and log2-transformed fold change ≥ 1.2 or ≤ −1.2 were considered as significantly differentially abundant. To define significantly differentially abundant taxa and GOs, aggregated iBAQ values of proteins of the same species or GO IDs were compared between women with low, medium, and high inflammation using the limma R package. The aggregation was performed separately for host and microbial proteins. The p values along with GO IDs were uploaded to REVIGO (http://revigo.irb.hr/) to visualize the top 50 biological processes with the lowest p values.

The effects of BV, STIs, and chemokine and proinflammatory cytokine profiles on the metaproteome were investigated by PCA using the mixOmics R package [37]. The ComplexHeatmap package [38] was used to generate heatmaps and cluster the samples and metaproteomes based on the hierarchical and K-means clustering methods. Additionally, R packages EnhancedVolcano, weighted correlation network analysis (WGCNA) [39], and FlipPlots were applied for plotting volcano plots, co-correlation analysis, and Sankey diagrams, respectively. Basic R functions and the gglplot2 R package were used for data manipulation, transformation, normalization, and generation of graphics. To identify the key factors distinguishing women defined as having low, medium, and high inflammation in their FGTs, we used random forest analysis using the R package randomForest [40] using the following settings: (i) the type of random forest was classification, (ii) the number of trees was 2000, and (iii) the number of variables evaluated at each split was one-third of the total number of variables [proteins/molecular function GO IDs/16S rRNA gene-based operational taxonomic units (OTUs)]. To compare the accuracy of the protein-, function- and species-based models, we iterated each random forest model 100 times for each of the input datasets separately. The distributions of the OOBs for 100 models for each data type were then compared using t tests. For the WGCNA analysis, a microbial-host weighted co-correlation network was built using iBAQ values for all proteins and modules were identified using dynamic tree cut. Correlations between proteins and modules were then determined and the top taxa and biological processes assigned to each of the proteins identified. Spearman’s rank correlation was used to determine correlation coefficients between individual cytokines/chemokines and module eigengenes (the first principal component of the expression matrix of the corresponding module) for all samples. Average Spearman’s correlation coefficients for all pro-inflammatory cytokines and chemokines were then determined. Similarly, Spearman’s correlation coefficients were calculated between module eigengenes and Nugent scores and STIs (including C. trachomatis, N. gonorrhoeae, T. vaginalis, M. genitalium, and active HSV-2).

16S rRNA gene sequencing and analysis

Bacterial 16S rRNA gene sequencing and analysis was conducted as previously described [27]. Briefly, the V4 hypervariable region of the 16S rRNA gene was amplified using modified universal primers [41]. Duplicate samples were pooled, purified using Agencourt Ampure XP beads (Beckman Coulter, CA, USA), and quantified using the Qubit dsDNA HS Assay (Life Technologies, CA, USA). Illumina sequencing adapters and dual-index barcodes were added to the purified amplicon products using limited cycle PCR and the Nextera XT Index Kit (Illumina, CA, USA). Amplicons from 96 samples and controls were pooled in equimolar amounts, and the resultant libraries were purified by gel extraction and quantified (Qiagen, Germany). The libraries were sequenced on an Illumina MiSeq platform (300-bp paired-end with v3 chemistry). Following de-multiplexing, raw reads were preprocessed, merged, and filtered using DADA2 [42]. Primer sequences were removed using a custom python script and the reads truncated at 250 bp. Taxonomic annotation was based on a customized version of SILVA reference database. Only samples with ≥ 2000 reads were included in further analyses. The representative sequences for each amplicon sequence variant were BLAST searched against NCBI non-redundant protein and 16S rRNA gene sequence databases for further validation, as well as enrichment of the taxonomic classifications. Microbial functional predictions were performed using Piphillin server [43] against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.
KSFM were used as a positive control [25]. Production of IL-6, IL-8, IL-1β, IP-10, MIP-3α, MIP-1α, MIP-1β, and IL-1RA were measured using a Magnetic Luminex Screening Assay kit (R&D Systems, MN, USA) and a Bio-Plex™ Suspension Array Reader (Bio-Rad Laboratories Inc®, CA, USA). VK2 cell viability following bacterial stimulation was confirmed using the Trypan blue exclusion assay. PCA was used to compare the overall inflammatory profiles of the isolates and to select the 22 most inflammatory and 22 least inflammatory isolates for characterization using proteomics.

Characterization of Lactobacillus protein expression in vitro using mass spectrometry

Forty-four Lactobacillus isolates were adjusted to 4.18 × 10⁶ CFU/ml in MRS and incubated for 24 h under anaerobic conditions. Following incubation, the cultures were centrifuged and the pellets washed 3x with PBS. Protein was extracted by resuspending the pellets in 100 mM triethylammonium bicarbonate (TEAB; Sigma-Aldrich, MO, USA) 4% sodium dodecyl sulfate (SDS; Sigma-Aldrich, MO, USA). Samples were sonicated in a sonication water bath and subsequently incubated at 95°C for 10 min. Nucleic acids were degraded using benzonase nuclease (Sigma-Aldrich, MO, USA), and samples were clarified by centrifugation at 10 000×g for 10 min. Quantification was performed using the Quanti-Pro BCA assay kit (Sigma-Aldrich, MO, USA). HILIC beads (ReSyn Biosciences, South Africa) were washed with 250 µl wash buffer [15% ACN, 100 mM Ammonium acetate (Sigma-Aldrich, MO, USA) pH 4.5]. The beads were then resuspended in loading buffer (30% ACN, 200 mM ammonium acetate pH 4.5). A total of 50 µg of protein from each sample was transferred to a protein LoBind plate (Merck, NJ, USA). Protein was reduced with tris (2-carboxyethyl) phosphine (Sigma-Aldrich, MO, USA) and alkylated with methylmethyleneisothiophosphonate (MMTS; Sigma-Aldrich, MO, USA). HILIC magnetic beads were added at an equal volume to that of the sample and a ratio of 5:1 total protein and incubated on the shaker at 900 rpm for 30 min. After binding, the beads were washed four times with 95% ACN. Protein was digested by incubation with trypsin for 4 h and the supernatant containing peptides was removed and dried down in a vacuum centrifuge. LC-MS/MS analysis was conducted with a Q-Exactive quadrupole-Orbitrap MS as described above. Raw files were processed with MaxQuant version 1.5.7.4 against a database including the Lactobacillus genus and common contaminants. Log₁₀-transformed iBAQ intensities were compared using the Mann-Whitney U test, and p values were adjusted for multiple comparisons using an FDR step-down procedure.

FGT-DB application

We have developed FGT-DB (Female Genital Tract Metaproteomics), an online application for exploring and mining the FGT microbial composition and function of South African women at high risk of HIV infection. The app is available at http://fgtdb.org/ and allows users to repeat the analyses presented here, as well as perform additional analyses and investigation using the metaproteomic data generated in this study.
Additional file 2: Table S2. List of protein BLAST results against UniProt and NCBI nr databases (IDb). Liquid chromatography-tandem mass spectrometry was used to evaluate the metaproteome in lateral vaginal wall swabs from 113 women from Cape Town, South Africa. Primary assigned taxa, obtained based on MaxQuant version 1.5.7.4, are presented in the second column. The curated taxa (third column) were obtained after considering the top similar and frequent BLAST hits for each protein. Ids: UniProt Identity codes. Table S3. List of differentially abundant taxa distinguishing women with low and high inflammation. Liquid chromatography-tandem mass spectrometry was used to evaluate the metaproteome in lateral vaginal wall swabs from 113 women from Cape Town, South Africa. The false discovery rate (FDR) corrected p-values were obtained based on the aggregated log2-transformed intensity-based absolute quantification of all proteins assigned to the same taxon applying the moderated t-test (limma R package), after adjusting for confounding variables including age, contraceptives, prostate specific antigen (PSA) and sexually transmitted infections (STIs). Taxa with FDR adjusted p-values \( \geq 0.05 \) and log2-transformed fold change \( \geq 1.2 \) or \( \leq -1.2 \) were considered as significantly abundant MFs. (PDF 6091 kb)

Table of differentially abundant proteins distinguishing women with low and high inflammation. Liquid chromatography-tandem mass spectrometry was used to evaluate the metaproteome in lateral vaginal wall swabs from 113 women from Cape Town, South Africa. The false discovery rate (FDR) corrected p-values were obtained based on the aggregated log2-transformed intensity-based absolute quantification of each protein applying the moderated t-test (limma R package), after adjusting for confounding variables including age, contraceptives, prostate specific antigen (PSA) and sexually transmitted infections (STIs). Proteins with FDR adjusted p-values \( \leq 0.05 \) and log2-transformed fold change \( \geq 1.2 \) or \( \leq -1.2 \) were considered as differentially abundant proteins. Ids: UniProt Identity codes. Table S5. List of differentially abundant molecular function (MF) gene ontology terms distinguishing women with low and high inflammation. Liquid chromatography-tandem mass spectrometry was used to evaluate the metaproteome in lateral vaginal wall swabs from 113 women from Cape Town, South Africa. The false discovery rate (FDR) corrected p-values were obtained based on the aggregated log2-transformed intensity-based absolute quantification of all proteins assigned to the same MF term applying the moderated t-test (limma R package), after adjusting for confounding variables including age, contraceptives, prostate specific antigen (PSA) and sexually transmitted infections (STIs). Taxa with FDR adjusted p-values \( \leq 0.05 \) and log2-transformed fold change \( \geq 1.2 \) or \( \leq -1.2 \) were considered as significantly abundant MFs. (PDF 6091 kb)
39. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
40. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2:18–22.
41. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, et al. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. MBio. 2014;5:e01283–14.
42. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581.
43. Narayan NR, Weinmaier T, Laserna-Mendieta EJ, Claesson MJ, Shanahan F, Dabagh K, et al. Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences. BMC Genomics. 2020;21:1–12.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.