N-glycosylation of cervicovaginal fluid reflects microbial community, immune activity, and pregnancy status

Gang Wu1,2,6, Paola Grassi1,2,6, David A. MacIntyre2,3, Belen Gimeno Molina2,3, Lynne Sykes2,3,4, Samit Kundu2,3, Cheng-Te Hsiao5, Kay-Hooi Khoo5, Phillip R. Bennett2,3, Anne Dell1,2* & Stuart M. Haslam1,2*

Human cervicovaginal fluid (CVF) is a complex, functionally important and glycan rich biological fluid, fundamental in mediating physiological events associated with reproductive health. Using a comprehensive glycomic strategy we reveal an extremely rich and complex N-glycome in CVF of pregnant and non-pregnant women, abundant in paucimannose and high mannose glycans, complex glycans with 2–4 N-Acetyllactosamine (LacNAc) antennae, and Poly-LacNAc glycans decorated with fucosylation and sialylation. N-glycosylation profiles were observed to differ in relation to pregnancy status, microbial composition, immune activation, and pregnancy outcome. Compared to CVF from women experiencing term birth, CVF from women who subsequently experienced preterm birth showed lower sialylation, which correlated to the presence of a diverse microbiome, and higher fucosylation, which correlated positively to pro-inflammatory cytokine concentration. This study is the first step towards better understanding the role of cervicovaginal glycans in reproductive health, their contribution to the mechanism of microbial driven preterm birth, and their potential for preventative therapy.

All cells have a sugar “coat” consisting of diverse glycans, which function as primary molecules for cellular recognition, adhesion, signalling and host–pathogen/host commensal interactions1–4. Human N-glycan biosynthesis starts with a precursor structure, from which high mannose, hybrid and complex type glycans are synthesized5. The antennae on complex and hybrid glycans carry important sequences, such as Lewis antigens, Sialyl Lewis antigens, ABO blood groups and poly LacNAc, which are recognised by glycan binding proteins (GBP), which are also known as lectins6,7. The glycans of glycoproteins expressed along the human female reproductive tract are of fundamental importance at all stages of human pregnancy, from conception and implantation to delivery2. In addition to acting as potential ligands for GBP of the host immune system, they can serve as adhesion and attachment points for both pathogenic and commensal microbes. Bacterial vaginosis (BV) is a polymicrobial imbalance of the vaginal microbiota, which is associated with development of precancerous cervical lesions, pelvic inflammatory disease, endometritis, tubal infertility and preterm birth8–11. The BV-associated bacteria Gardnerella vaginalis and Prevotella bivia produce extracellular sialidases to hydrolyse mucosal sialoglycans and transport the released sialic acid inside the bacterium cells for catabolism12,13, which may benefit their survival under some nutrient deprived conditions. Recent proteome-wide prediction studies of bacteria have identified > 100,000 putative bacterial GBP14,15. Genome screening found that vaginal bacterial species associated with infection and inflammation produce a larger number of GBP than commensals, potentially allowing them to bind a wider range of glycans in the vagina14. Moreover, the number of predicted bacterial lectins and their specificities correlated with pathogenicity. Secretor status, related to the expression of the FUT2 gene, is responsible for the synthesis of the H-antigen glycan structure in the mucus and also provides the precursor for mucosal blood group A and B synthesis. Approximately 20% of the human population presents a mutation in

1Department of Life Sciences, Imperial College London, London, UK. 2March of Dimes Prematurity Research Centre at Imperial College London, London, UK. 3Institute of Reproductive and Developmental Biology, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, UK. 4The Parasol Foundation Centre for Women’s Health and Cancer Research, St Mary’s Hospital, London W1 2NY, UK. 5Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan. 6These authors contributed equally: Gang Wu and Paola Grassi. *email: a.dell@imperial.ac.uk; s.haslam@imperial.ac.uk
FUT2, which abolishes α(1,2)fucosyltransferase activity. This ‘nonsecretor’ phenotype is associated with a 2.4 to 4.4 fold increased relative risk for recurrent vaginitis by *C. albicans*16, and maternal lack of H-antigen production was found to be a risk factor for preterm birth17. Moreover, non-secretors with *Lactobacillus* depleted vaginal microbiota in early pregnancy have significantly shorter pregnancies compared to women with *Lactobacillus* dominated composition, a relationship not seen in women who are secretors18.

Human physiological parturition is a pro-inflammatory process39. Histological studies have found leukocytes, especially macrophages and neutrophils, infiltrate the human myometrium, fetal membrane and the cervix, around the time of parturition20,21, which is accompanied by increased levels of pro-inflammatory cytokines, such as IL-1β, IL-6, IL-8 and TNF-α. Preterm labour is associated with early activation of inflammation at the cervical-vaginal and maternal–fetal interface, and is commonly thought to be due to the presence of pathogenic microbes22,23. Neutrophils are the key immune cells recruited to fetal membranes as a result of microbial driven inflammation33, and are likely to be key immune mediators at the cervical-vaginal interface25. Infection is thought to contribute to at least one third of preterm birth cases1,39, and there is mounting evidence for the role of host-vaginal microbial interactions in these processes20,22,23,25. The cervix functions as both a mechanical and immunological barrier to ascending infection during pregnancy. This is in part mediated by cervical secretion of highly glycosylated glycoproteins, which are a major constituent of cervical mucus28,29. It is estimated that up to 80% of the weight of cervical mucus can be attributed to glycans30,31. Cervicovaginal fluid (CVF) thus represents a heterogeneous mixture of endocervical and vaginal secretions, as well as epithelial cells, leukocytes and antibodies, all of which are highly decorated by complex glycan structures. In other mucosal niches, glycan mediation of immune response is well characterised and includes crosstalk between glycans and immune receptors that regulate leukocyte migration, antigen presentation, immune activation and antibody responses3,32–34.

Proteomic analyses of CVF have identified more than 1200 proteins in the CVF35, many of which can be functionally associated with immune responses, infection, and pregnancy complications including biochemical processes involved in the onset of preterm labour27,36–45. Recently a desorption electrospray ionization mass spectrometry (DESI-MS) based study, which allows ionisation of mucosal biomass directly from a standard clinical swab, demonstrated changes in the mucosal metabolome associated with pregnancy, microbiota composition and immune activation that associates with preterm birth risk46.

Despite their importance in mediating various physiological events associated with reproductive health, particularly during pregnancy, to our knowledge, detailed profiling of CVF glycans has not been reported. In this study, we used a comprehensive glycomic strategy to characterize the N-glycans in the CVF of pregnant and non-pregnant women. We demonstrate that glycan structural features such as fucosylation and sialylation are associated with pregnancy status, microbial community state, immune activity, and pregnancy outcome.

**Results**

**N-glycan structural profiling.** CVF samples from a total of 10 donors with diverse ethnicities were collected for analysis. 4 of the donors were non-pregnant women, while 6 of the 10 donors were pregnant women defined at high risk of spontaneous preterm delivery (sPTB). Of these, 3 delivered at term, while 3 delivered preterm (<37 weeks of pregnancy) (Table 1, Supplementary Table S1).

A detailed MALDI-MS based glycomic characterisation of permethylated N-glycans was undertaken. This produces [M + Na]^+ molecular ions. An overall relative quantitation of paucimannose, high mannose, hybrid and complex glycans is shown in Supplementary Fig. S1. More detailed structural characterization is illustrated in Fig. 1. Paucimannose glycans were identified in the low mass range at m/z 1141, 1171, 1345 and 1375, (ManαβGlcNAc3Fucα1). The glycans in the mid mass range were dominated by high mannose glycans (m/z 1579–2396, ManαβGlcNAc3) and bi-antennary complex glycans with fucosylation and/or sialylation (m/z 1835–3211, NeuAcα2Galβ1ManαβGlcNAc3Fucα1). The glycans at higher mass range contained multiple LacNAc units (Galβ1GlcNAc) modified by different numbers of fucose and sialic acid (m/z 3402–4650, NeuAcα2Galβ1ManαβGlcNAc3Fucα1). The largest complex N-glycan observed was at m/z 6259 corresponding to a composition of Galβ1ManαβGlcNAc3Fucα1 (Supplementary Table S2). More detailed N-glycan structural analysis was achieved by MS/MS analysis of selected molecular ions. This proved that the mono-fucosylated glycans can be both core or antenna fucosylated (Supplementary Fig. S2a), and that peaks with composition consistent with

| Sample code | Pregnancy outcome | CST | Ethnicity | Gestation (weeks + days) | ABO blood group |
|-------------|-------------------|-----|-----------|-------------------------|-----------------|
| NP1         | Non-pregnant      | Not available | Caucasian | 0 + 0                   | O               |
| NP2         | Non-pregnant      | Not available | South Asian | 0 + 0                   | A               |
| NP3         | Non-pregnant      | Not available | African    | 0 + 0                   | A               |
| NP4         | Non-pregnant      | Not available | Caucasian | 0 + 0                   | A               |
| P1          | Term              | I    | Caucasian | 39 + 0                  | O               |
| P2          | Term              | I    | African   | 39 + 3                  | B               |
| P3          | Term              | IV-B | African   | 39 + 3                  | O               |
| P4          | Preterm           | I    | South Asian | 27 + 3                  | B               |
| P5          | Preterm           | IV-B | African   | 31 + 3                  | B               |
| P6          | Preterm           | IV-B | South Asian | 22 + 4                  | B               |

**Table 1.** Summary of CVF samples collected for glycomic analysis.
multiple LacNAc units are a mixture of structural isomers with varying numbers and lengths of antennae. Complex glycans with 2–4 LacNAc units were identified in the medium mass range and glycans with poly LacNAc units in the high mass range (Fig. 1). Glycans above m/z 5000 included high numbers of LacNAc units exhibiting diverse modifications by Fuc and NeuAc (Supplementary Table S2).

The MS/MS analysis reveals that structures with extended antennae are often favoured over branched structures (Supplementary Fig. S2b). These extended antennae were decorated by several fucose residues producing diverse poly Lewis antigens. LacdiNAc (Gal[α1-3GalNAcβ1-4GlcNAc]) containing structures were identified at m/z 2285, 2326, 2500 and 2674, which were supported by MS/MS fragmentation analysis (Supplementary Fig. S2c). Interestingly, sample P5, from a donor with CST IV-B who delivered preterm at 31 + 3 weeks, showed a distinct pattern of glycans rich in fucosylated glycans and mono-fucosylated glycans, but lower levels of bi-, tri- and tetra-fucosylated glycans compared to CST I-A/B samples. Our results suggest that the presence of a high relative abundance of Lactobacillus spp. is associated with a higher percentage of sialylation, bi-, tri- and tetra-fucosylated glycans, and lower levels of non- or mono-fucosylated glycans.

A distinct N-glycan composition was also observed in CVF from non-pregnant donors (Fig. 1e, Supplementary Fig. S4). Compared to pregnant women, spectra obtained from non-pregnant donors are dominated in the lower mass range by high mannose N-glycans (m/z 1579–2396, Manα(1–6)GlcNAcβ), while the mid mass range is dominated by biantennary complex poly-fucosylated structures at m/z 2592, 2766 and 2940 (GalαManαGlcNAcFuc3-5). The higher mass range is dominated by large polyLacNAc, heavily fucosylated glycans such as structures at m/z 3215, 3563, 3737, 4012, 4186, 4361, 4810 and 4984 (GalαManαGlcNAcFuc4-5). Complex biantennary, triantennary and tetranantennary glycans capped by one to four sialic acids, were also detected, (eg, m/z 2605, 2966, 3211, 3776, 4124, 4473, 4761) but at lower abundance than exclusively fucosylated glycans.

The complexity of CVF N-glycans with their multiple fucosylation and sialylation led us to apply additional complementary analytical strategies to facilitate deeper characterization of CVF N-glycan structures. Glyco-epitope centric N-glycan analysis using an Orbitrap mass spectrometer confirmed the presence of diverse glycotopes in a selected sample P3 (Fig. 2). MS2 fragments of N-glycans detected LacdiNAc structures at m/z 505 and 679, terminal Lewis antigens at m/z 638 and 812, internal Lewis antigens at m/z 624, and Sialyl Lewis antigens at m/z 999. In addition, poly Lewis antigens were observed at m/z 1261 and m/z 1435. The MS2 data also showed terminal GlcNAc or GalNAc at m/z 260, predicted to arise from bisecting glycans, glycans with truncated antennae, or LacdiNAc structures. The MS2 peak at m/z 812 was selected for further fragmentation to obtain MS3 data, which mainly detected the Lewis Y antigen. However, the MS3 analysis of the MS2 peak at m/z 638 revealed a mixture of Lewis X, Lewis A and the blood group H antigen, again highlighting the structural complexity of CVF N-glycans.
Glycans in the lower mass range, which could be a sign of reduced glycan maturation and differentiation or a transition to the higher mass range. However, spectra of non-pregnant women CVF are clearly dominated by high mannose glycans in the low mass range, biantennary complex poly-fucosylated structures in the mid mass range, and large PolyLacNAc glycans in the high mass range. These findings suggest a biological role of the elongated antennae protruding from the mucin protein backbone. The structure of the glycan code is represented in the corresponding coloured rectangle in panel (f). Main structures are depicted. Assignments are based on composition, tandem MS and knowledge of biosynthetic pathways. All molecular ions are [M + Na]+. Residues above a bracket have not had their location unequivocally defined.

**Correlation of fucosylation and sialylation to pro-inflammatory cytokines.** Strong correlation was found when fucosylation was mapped to levels of IL-1β and IL-18 (Figs. 8 and 9). The proportion of non-fucosylated glycans (Fig. 8a) and the proportion of mono-fucosylated glycans (Fig. 8b) to total glycans were positively correlated with cytokine concentration, while the proportion of poly-fucosylated glycans to total glycans (Fig. 8c) and the proportion of highly-fucosylated glycans to poly-fucosylated glycans (Fig. 8d) were negatively correlated with the cytokine concentrations. Similar correlations were observed for glycans with 3 LacNAc units (Fig. 9). No statistically significant correlations were seen with IL-8 and IL-6, however similar trends were observed (Supplementary Figs. S6 and S7). We also investigated levels of glycan sialylation and pro-inflammatory cytokines, however, only weak correlations were observed (Supplementary Figs. S1–S11).

**Discussion**

Despite the importance of human CVF in shaping microbiota-host responses in the lower reproductive tract, to our knowledge, a detailed profiling of CVF N-glycans has not been reported. Our study characterises CVF as a glycans-rich environment with distinct and different N-glycosylation in relation to pregnancy status, microbial composition, immune activation, and pregnancy outcome. Our findings are based on a small cohort at present. This is due to a combination of factors including the collection device being more invasive than a simple swab and the labour intensive and high cost of sample processing. However, some initial trends have already been consistently observed. We also acknowledge that factors such as the stage of menstrual cycle or pregnancy, age, cervical length, blood group and ethnicity could influence the CVF glycosylation profile. We aim to study these factors with larger cohorts. The N-glycome of CVF is highly complex and abundant in paucimannose glycans and high mannose glycans in the low mass range. Complex glycans with 2–4 antennae in the medium mass range and glycans with poly LacNAc units in the high mass range are readily detectable. Antennae structures of detected N-glycans included terminal fucosylation ofmono-fucosylated glycans, glycans with LacdiNAc structures, as well as sialylation and fucosylation patterns of extended poly LacNAc structures. Identification of a higher prevalence of glycan isomers carrying extended poly LacNAc chains compared to multi-antennary structures, suggests a biological role of the elongated antennae protruding from the mucin protein backbone. While extended poly-N-acetyllactosamine glycans can serve as spacers and carriers of additional glycoconjugates through their modification with fucose and sialic acid, poly LacNAc itself has been identified as a recognition motif of galectins, an important class of mammalian GBP involved in regulations of immune cell activities and microbial recognition as part of the innate immune system. N-Glycan structures carrying extended poly LacNAc chains have recently been reported as a characteristic feature of neutrophil granules, which suggests these structures in CVF might partially be due to leukocyte infiltration. High mannose glycans can be recognized by DC-SIGN, the mannose receptor and other mannose binding GBP, which have been found to be involved in inflammation and infection.

Modifications of genital tract epithelia, mucus, and immune status occur throughout the menstrual cycle to accommodate conception and implantation. At the time of conception, Sialyl Lewis X glycan epitopes on the human oocyte facilitates sperm–egg binding, and glycans have also been implicated in the mediation of implantation and immune homeostasis at the maternal–fetal interface. Once implantation is successful, the cervicovaginal interface undergoes further modification, with the primary aim of forming a cervical barrier to infectious microbes to protect the developing fetus. Our study shows that this physiological switch to a pregnancy phenotype also involves marked changes to cervicovaginal glycoprofiles. In both non-pregnant and pregnant women, N-glycans were found to be a mix of paucimannose and high-mannose structures in the lower mass range, biantennary complex poly-fucosylated structures in the mid mass range, and large PolyLacNAc glycans in the high mass range. However, spectra of non-pregnant women CVF are clearly dominated by high mannose glycans in the lower mass range, which could be a sign of reduced glycan maturation and differentiation or a different microbiota composition compared to pregnant women. Oligosaccharides carrying terminal mannose residues are the main host receptors for bacteria such as *E. coli*. A loss of high mannose residues on glycoproteins of the CVL from women with BV was observed by Wang et al., suggesting that native high mannose residues found on glycoproteins in fluids of the lower reproductive tract may act as natural inhibitors of pathogenic interactions, helping to prevent bacterial adhesion in women with normal microflora.
Figure 2. Glycotope centric analysis of N-glycans from sample P3. All N-glycans were selected for CID fragmentation to produce MS2 ions. The summed intensity for each of the detected diagnostic MS2 ions normalized to their total was plotted to provide an overall assessment of the relative abundance of various glycotopes. The glycotopes represented by MS2 ions at m/z 638 and m/z 812 were further selected for additional stage of fragmentation (MS3) to identify their respective isomeric constituents, as shown in the insets.

Figure 3. Overall view of fucosylation and sialylation levels of N-glycans with 2 LacNAc units (a,b) and 3 LacNAc units (c,d). The glycans were subgrouped according to the number of Fuc or NeuAc they have. The summed intensity of glycans in a subgroup divided by the total glycan intensity was calculated as a relative intensity of the subgroup.
Glycans in the mid to high mass range of non-pregnant women show lower fucosylation and sialylation levels compared to pregnant women. This could be pointing again at a lower maturation and differentiation state of cervicovaginal proteins N-glycans and to a more quiescent immunological state which would allow easier transit of sperm and thus easier fertilisation.

We also demonstrated differences in glycan profiles between women who subsequently delivered preterm and women who delivered at term. The preterm samples showed higher levels of non-fucosylated glycans among glycans with 2 or 3 LacNAc units and lower sialylation among poly-fucosylated glycans. Desialylation of glycoproteins and glycolipids by sialidase leads to loss of or exposure of new glycan epitopes and interference with host immune recognition and binding of microbes. It is feasible that this could contribute to the mechanism of microbial driven preterm birth. In support of this, high sialidase concentrations have been associated with late miscarriage and preterm delivery. Further, we and others have reported an association between vaginal microbial composition and preterm birth risk. High diversity vaginal microbial composition (referred to as community state type IV, CST IV), or dominance of the niche by L. iners, is associated with higher rates of preterm birth, whereas L. crispatus (CST I) dominance is protective. In this study N-glycans with 2 and 3 LacNAc units isolated from CVF of women with CST IV have lower levels of both poly and highly-fucosylated glycans and sialylated glycans compared to women with CST I.

Fucosylation and sialylation on LacNAc units can potentially produce different glycotopes, such as Lewis antigens, Sialyl Lewis antigens and Poly Lewis antigens. Several studies have shown how glycans and their epitopes play a role in infective conditions such as HIV and Candida albicans. G. vaginalis is one of the main bacteria associated with bacterial vaginosis and CST IV, which is characterised by depletion of Lactobacillus species and polymicrobial overgrowth of anaerobes. BV is also associated with increased vaginal concentrations of glycosidases such as sialidase, α-galactosidase, β-galactosidase and α-glucosidase and decreased binding of lectins to both high mannose and α-2, 6 sialic acid. This observation suggests that bacteria sialidases and host sialoglycan receptors can be linkage specific, as also reported by other studies. Findings from previous studies suggest that changes in glycosidases are accompanied by changes in glycosylation patterns in the vaginal fluid. On the other hand, a recent discovery found that Lactobacillus spp. Can create a protective micro-ecological environment through regulating the core fucosylation of vaginal epithelial cells. Although glycosylation of the CVF could be influenced by hormone levels, it is thought that microbiota have a greater influence on CVF glycosylation than hormonal status.

Similarly to CVF, the human gastrointestinal (GI) mucosa consists of extensive glycoproteins overlaying epithelial cells, interspersed by immune cells, which form a protective physical barrier, and serve as first line of defence against microorganisms and harmful substances. Hydrolases, including exo- and endoglycosidases, are encoded in the genomes of mucin-degrading gut bacteria allowing them to metabolize sugars, which are then exploited as nutrient source. The mucins of the GI tract mucus layer also provide ligands for the attachment of bacteria and may facilitate invasion; moreover, mucus-binding proteins such as adhesins and GBP have been described in many lactic acid bacilli. The human gut microbiome and its role in both health and disease has been extensively studied and its involvement in human metabolism, nutrition, physiology, and immune function is now well established. It is highly probable that similar interactions exist in the mucosa of the female lower reproductive tract, creating a tight interconnection between glycans, microorganisms colonising the female reproductive tract and cervicovaginal immune homeostasis.
It is well established that local inflammation is a risk factor for preterm birth. However, only recently are we beginning to understand the relationship between inflammation, specific microbial communities, and preterm birth. Here we observed that non-fucosylated and monofucosylated glycans positively correlated with pro-inflammatory cytokines, whereas poly and high-fucosylated glycans are negatively correlated, highlighting the important role of N-glycosylation in maintaining immune homeostasis in the female reproductive tract. An abundance of paucimannose glycans was also detected in the low mass range of CVF, which could originate from infiltrating neutrophils, monocytes and macrophages. These glycans could be derived from a recently described noncanonical truncation N-glycosylation pathway in human neutrophils. Paucimannose glycans are found on proteins in azurophilic granules of neutrophils, which can be selectively secreted upon pathogen stimulation, and their glycosylation status is involved in modulating multiple immune functions central to inflammation and infection. The unusually high abundance of paucimannose structures found in CVF is potentially indicative of neutrophil invasion in the lower female reproductive tract, which could be used as a defence mechanism against invading pathogens. It is plausible that glycan-GBP interactions between pathogenic microbes and neutrophils leads to neutrophil degranulation, an abundance of paucimannoses, pro-inflammatory cytokine release and subsequent preterm labour. Another possible explanation for the paucimannose glycans is that they are the result of degradation by microbial glycohdrolytic enzymes.

It should be noted that CVF is also heavily O-glycosylated and we are undertaking similar analytic approaches to characterise these glycan structures.

In summary, this study provides evidence that cervicovaginal glycans are associated with microbial composition and immune response in pregnancy where they may influence clinical outcome. Modulation of cervicovaginal glycans could thus represent a novel therapeutic strategy to prevent microbial driven preterm birth.
Methods

Patient recruitment and sampling. The study was conducted with approval of the NHS National Research Ethics Service (NRES) Committees London—Stanmore (REC 14/LO/0328), and in accordance with relevant guidelines, regulations, and the Declaration of Helsinki. All methods were performed in accordance with the relevant guidelines and regulations. All patients and non-pregnant women provided written informed consent. Recruitment and sampling were performed at Queen Charlotte’s and Chelsea Hospital, Imperial Col-
lege Healthcare NHS Trust, London, UK. Non pregnant women were eligible if they were of reproductive age and aged 18 or over. Pregnant women at risk of preterm birth were eligible. Risk factors included having a short or open cervix, a previous preterm delivery or previous cervical treatment. Exclusion criteria included women under 18 years of age, those who had sexual intercourse within 72 h of sampling, vaginal bleeding in the preceding week, HIV or Hepatitis C positive status. Detailed maternal clinical metadata and birth outcome data was collected for all pregnant participants. CVF was sampled using the BBL™ CultureSwab™ MAXV liquid Amies swabs (Becton, Dickinson and Company, Oxford UK) for assessing microbial composition. CVF was then collected using a menstrual cup (Softdisc™, The Flex Company, USA) by placing it against the cervix for 20 min. After removal, material from both sides of the cup was retrieved by repeated pipetting of phosphate buffer saline (PBS) over each side leading to resuspension of material in a 1:5 weight: volume ratio within 30 min of collection. The suspension was distributed into separate aliquots to prevent unnecessary freeze thaw cycles. Aliquot one was reserved for glycomic profiling. Aliquot two was centrifuged (500×g, 10 min, 4 °C), and the supernatant was reserved for immune profiling. Samples were stored at – 80 °C until analysis.

**N-Glycan profiling of CVF samples.** Methanol, acetonitrile, ammonia, chloroform, DMSO, propan-1-ol, sodium hydroxide, acetic acid were from Romil (Cambridge, UK). Idoacetic acid, sodium chloride, iodometh-
ane, ammonium bicarbonate, EDTA, trypsin and Tris were from Merck (Poole, UK). PNGase F (cloned from Flavobacterium meningosepticum and expressed by E. coli), CHAPS and DTT were from Roche Applied Science (East Sussex, UK). 8 M guanidine hydrochloride (GuHCl) and Slide-A-Lyzer™ G2 Dialysis Cassettes, 3.5K MWCO were from Thermo scientific (Loughborough, UK).

Glycomic sample processing was done following the protocol detailed previously. Briefly, CVF samples from aliquot one were sonicated in 25 mM Tris, 150 mM NaCl, 5 mM EDTA, and 1% CHAPS, pH 7.4, dialysed in dialysis cassettes, reduced by DTT, carboxymethylated by IAA, and digested by trypsin. N-glycans were released by PNGase F and were permethylated. The permethylated glycans were cleaned by C18 cartridges and freeze dried before mass spectrometry analysis. Glycan profiling was done on AB Sciex 4800 MALDI-TOF/TOF mass spectrometer. The methylated glycans were dissolved in 10 ul methanol. One 1ul of sample was mixed with 1 ul of 10 mg/ml DABP matrix in 75% ACN. The mixture was spotted on a MALDI plate for MALDI-TOF–MS analysis. The data were analysed using Data Explorer™ version 4.6 from AB Sciex, Glycoworkbench92 and MALDIquant93. The glycomic data were annotated based on monosaccharide composition derived from the molecular ion m/z value, knowledge of N-glycan biosynthetic pathways, the isotopic peak cluster patterns, the glycosylation patterns in the low and medium mass range, and MS/MS derived fragmentation. For glycans with overlapped isotopic peak clusters, the dominant one was annotated.

The statistical analysis of the glycans was done using SQLite3 and R 3.6.3 on a Linux Ubuntu 20.04.4 LTS platform. The scripts are available on GitHub (Link: https://github.com/gw110/CVF-glycosylation.git). In order to have the least interference of overlapped isotopic peaks, Glycans with 2 or 3 LacNAc units were selected for quantitative analysis to assure high accuracy of data annotation and quantitation. The relative intensities of different glycan subgroups were calculated using SQLite3. The results were exported and visualized using ggplot2 in R 3.6.3. The correlation of glycan relative intensities with cytokines was done using the linear regression model. Pearson product-moment correlation coefficient and p value were calculated using the cor.test function in R. Box plot and two tailed Student’s t test was used to compare glycosylation between sample groups.

Glycans were further analysed following a glyco-centric glycomic analysis based on a nanolC-MS²-product dependent-MS³ data acquisition workflow, previously established using the advanced Orbitrap Fusion Tribrid MS platform. In essence, this takes advantage of the high acquisition speed, high mass accuracy and high sensitivity of the MS instrument to perform a comprehensive MS² mapping of the terminal glycocone based on preferred cleavage at the HexNAc of permethylated glycans. The respective intensity of each characteristic oxonium ion measured at 5 ppm mass accuracy or less can be summed from all glycan MS² spectral acquired throughout the LC–MS/MS run to provide a relative abundance index for the glycocone it represents. In cases where isomeric glycocones need to be differentiated, as for the Lewis and H epitopes in this work, the MS² ion can be coupled with a further MS³ event. The diagnostic MS³ ions resulting from specifically eliminating the C3-substituent of the HexNAc⁺ is sufficient to resolve the isomeric difference and their relative intensity can in turn provide a semi-quantitative estimation of the relative abundance of each isomeric constituent of that particular glycocone.

**Bacterial DNA extraction and metataxonomic profiling.** Extraction of DNA from CVF samples and sequencing of 16S rDNA hyper variable regions was performed as previously described. Briefly, V1-V2 hyper variable regions of bacterial 16S rRNA genes were amplified using a mixed forward primer set (28f.-YM) consisting of the following primers mixed at a 4:1:1:1 ratio; 28F-Borreella GAGTTTGATCCTGCGTAT; 28F-Chlorrelx GATTGTGCTTGTTGAC; 28F-Bifido GGTTGATCCTGCGTAT; 28F GATTTGATCTGCTGTC. The reverse primer consisted of; 388R TGC TGC CTC CCG TAG GAG T95. Amplified products were then pooled equimolar and each pool was size selected in two rounds using Agencourt AMPure XP (Beckman Coulter). The amplicon sequence variant (ASV) counts were calculated for each sample using the Qiime2 pipeline. DADA2 was used for denoising and taxonomic classification of sequences to species level was performed using the STIRRUPS reference database. Samples were classified into vaginal community state types (CSTs) using the VaginaL community state type Nearest Centrold classifier (VALENCA). This using this standardised approach, CST 1 communities are characterised by Lactobacillus crispatus dominance and can be further subdivided into subtypes CST 1-A (almost complete dominance by L. crispatus) and CST 1-B (less L. crispatus, but still the majority). CST IV communities have a low relative abundance of Lactobacillus spp. and include subtypes CST IV-A, CST IV-B, and CST IV-C, depending on the relative abundances of G. vaginalis, A. vaginae, and BVAB1. respectively.

**Immune profiling of CVF samples.** The stored supernatant from the CVF (aliquot two) was thawed on ice and used for Luminex® immunoassays to quantify cytokines by multiplexed bead-based immunoassays. The concentrations of 10 cytokines (IL-1β, IL-18, IL-6, IL-2, IL-4, IL-5, IL-10, IFN-γ, GM-CSF and TNF-α) were measured on a multiplex plate Human Premixed Multi-Analyte Kit (R&D Systems), following manufacturer’s instructions. IL-8 was measured on a single-plex Human Premixed Analyte Kit (R&D Systems), following a 10-fold dilution using Calibrator Diluent RD6-52. All standards and samples were run in duplicate. Concentrations of IL-2, IL-5, IFN-γ, and GM-CSF were undetectable in all samples. IL-4, IL-10 and TNF-α were only detectable in a proportion of samples.
Data availability
The R and SQLite3 scripts for glycan quantitative data analysis are available on GitHub (Link: https://github.com/gw110/CVF-glycosylation.git).

Received: 9 June 2022; Accepted: 15 September 2022
Published online: 10 October 2022

References
1. Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. Nat. Rev. Immunol. 7, 255–266 (2007).
2. Bloits, S. M. et al. Role of galectin-glycan circuits in reproduction: From healthy pregnancy to preterm birth (PTB). Semin. Immunopathol. 42, 469–486 (2020).
3. Vagios, S. & Mitchell, C. M. Mutual preservation: A review of interactions between cervicovaginal mucus and microbiota. J. Proteome Res. 14, 229–236 (1999).
4. Osman, L. et al. Leukocyte density and pro-inflammatory cytokine expression in human fetal membranes, decidua, cervix and myometrium before and during labour at term. Mol. Hum. Reprod. 9, 41–45 (2003).
5. Chan, D. et al. Microbial-driven preterm labour involves crosstalk between the innate and adaptive immune response. Nat. Commun. 13, 975 (2022).
6. Romero, R. et al. The role of inflammation and infection in preterm birth. Semin. Reprod. Med. 25, 21–39 (2007).
7. Tong, M. & Abrahams, V. M. Neutrophils in preterm birth: Friend or foe?. Front. Immunol. 13, 841723 (2022).
8. Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. Lancet Lond. Engl. 371, 75–84 (2008).
9. Kindlinger, L. M. et al. Relationship between vaginal microbial dysbiosis, inflammation, and pregnancy outcome in cervical cerclage. Sci. Transl. Med. 8, 35ra102 (2016).
10. Daunter, B. & Cousinmait, C. Cervical mucus: Its structure and possible biological functions. Eur. J. Obstet. Gynecol. Reprod. Biol. 10, 141–161 (1980).
11. Lacroux, G., Gouyard, V., Gottrand, F. & Desseyen, J.-L. The cervicovaginal mucus barrier. Int. J. Mol. Sci. 21, E8266 (2020).
12. Kim, Y. E., Kim, K., Oh, H. B., Lee, S. K. & Kang, D. Quantitative proteomic profiling of Cervicovaginal fluid from pregnant women with term and preterm birth. Proteome Sci. 19, 3 (2021).
13. Carson, D. D. et al. Mucin expression and function in the female reproductive tract. Hum. Reprod. Update 4, 459–464 (1998).
14. Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. Nat. Rev. Immunol. 8, 874–887 (2008).
15. Ruddle, F. P., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. Science 291, 2370–2376 (2001).
16. Zhou, J. Y. & Cobb, B. A. Glycans in immunologic health and disease. Annu. Rev. Immunol. 39, 511–536 (2021).
17. Dasiari, S. et al. Comprehensive proteomic analysis of human cervical-vaginal fluid. J. Proteome Res. 6, 1258–1267 (2008).
18. Moncla, B. J., Chappell, C. A., Debo, B. M. & Meyn, L. A. The effects of hormones and vaginal microflora on the glycome of the cervicovaginal fluid. PLoS ONE 11, e0158687 (2016).
19. Lockwood, C. J. et al. Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery. N. Engl. J. Med. 325, 669–674 (1991).
20. Conde-Agudelo, A. & Romero, R. Cervical phosphorylated insulin-like growth factor binding protein-1 test for the prediction of preterm birth: A systematic review and metaanalysis. Am. J. Obstet. Gynecol. 214, 57–73 (2016).
21. Melchior, J. C., Khalil, A., Wing, D., Schleusner, E. & Surbek, D. Prediction of preterm delivery in symptomatic women using PAMG-1, fetal fibronectin and pHGRFP-1 tests: Systematic review and meta-analysis. Ultrasound Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstetr. Gynecol. 52, 442–451 (2018).
22. Dos Santos, F., Daru, J., Rogozinska, E. & Cooper, N. A. M. Accuracy of fetal fibronectin for assessing preterm birth risk in asymptomatic pregnant women: A systematic review and meta-analysis. Acta Obstet. Gynecol. Scand. 97, 657–667 (2018).
44. Rizzo, G.
43. Rizzo, G.
45. Tanaka, Y.
46. Pruski, P.
47. Pang, P.-C.
48. Ugonotti, J.
49. Antonopoulos, A.
50. Drickamer, K. Two distinct classes of carbohydrate-recognition domains in animal lectins.
51. Vasta, G. R. Roles of galectins in infection.
52. Di Lella, S.
53. Josenhans, C., Müthing, J., Elling, L., Bartfeld, S. & Schmidt, H. How bacterial pathogens of the gastrointestinal tract use the
54. Cornick, S., Tawiah, A. & Chadee, K. Roles and regulation of the mucus barrier in the gut.
55. Firon, N., Ofek, I. & Sharon, N. Interaction of mannose-containing oligosaccharides with the fimbrial lectin of
56. Wang, L.
57. Kindinger, L. M.
58. Passaponti, S., Pavone, V., Cresti, L. & Ietta, F. The expression and role of glycans at the feto-maternal interface in humans.
59. Fettweis, J. M.
60. Lazar, V.
61. Ravcheev, D. A. & Thiele, I. Comparative genomic analysis of the human gut microbiome reveals a broad distribution of meta-
62. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
63. Josenhans, C., Müthing, J., Elling, L., Bartfeld, S. & Schmidt, H. How bacterial pathogens of the gastrointestinal tract use the
64. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
65. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
66. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome.
67. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
68. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
69. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
70. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
71. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
72. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
73. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
74. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
75. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
76. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
77. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
78. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
79. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
80. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
81. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
82. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
83. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
84. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
85. Ugonotti, J., Chatterjee, S. & Thaysen-Andersen, M. Structural and functional diversity of neutrophil glycosylation in innate immunity and related disorders. *Mol. Aspects Med.* 79, 100882 (2021).

86. Tjondro, H. C., Loke, I., Chatterjee, S. & Thaysen-Andersen, M. Human protein paucimannosylation: Cues from the eukaryotic kingdoms. *Biol. Rev. Camb. Philos. Soc.* 94, 2068–2100 (2019).

87. Loke, I., Østergaard, O., Heegaard, N. H. H., Packer, N. H. & Thaysen-Andersen, M. Paucimannose-rich N-glycosylation of spatiotemporally regulated human neutrophil elastase modulates its immune functions. *Mol. Cell. Proteomics* 16, 1507–1527 (2017).

88. Reiding, K. R., Lin, Y.-H., van Alphen, E. P. J., Meijer, A. B. & Heck, A. J. R. Neutrophil azurophilic granule glycoproteins are distinctively decorated by atypical pauc- and phosphomannose glycans. *Commun. Biol.* 4, 1012 (2021).

89. Thaysen-Andersen, M. et al. Human neutrophils secrete bioactive paucimannosidic proteins from azurophilic granules into pathogen-infected sputum. *J. Biol. Chem.* 290, 8789–8802 (2015).

90. Jang-Lee, J. et al. Glycomic profiling of cells and tissues by mass spectrometry: Fingerprinting and sequencing methodologies. *Methods Enzymol.* 415, 59–86 (2006).

91. North, S. J. et al. Mass spectrometric analysis of mutant mice. *Methods Enzymol.* 478, 27–77 (2010).

92. Ceroni, A. et al. GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* 7, 1650–1659 (2008).

93. Gibb, S. & Strimmer, K. MALDIquant: A versatile R package for the analysis of mass spectrometry data. *Bioinform. Oxf. Engl.* 28, 2267–2271 (2012).

94. Hsiao, C.-T. et al. Advancing a high throughput glyco-tope-centric glycomics workflow based on nanoLC-MS2-product dependent MS3 analysis of permethylated glycans. *Mol. Cell. Proteomics* MCP 16, 2268–2280 (2017).

95. Frank, J. A. et al. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470 (2008).

96. Kechin, A., Boyarskikh, U., Kel, A. & Filpenko, M. cutPrimers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* 24, 1138–1143 (2017).

97. Babraham Bioinformatics. FastQC A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

98. Bolyn, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857 (2019).

99. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583 (2016).

100. Fettweis, J. M. et al. Species-level classification of the vaginal microbiome. *BMC Genomics* 13(Suppl 8), S17 (2012).

**Acknowledgements**

We would like to thank all women who have participated in this study and members of the Women’s Health Research Centre who facilitated and coordinated study recruitment and sample collection. This work was funded by the March of Dimes European Preterm Birth Research Centre at Imperial College London and supported by the National Institute of Health Research (NIHR) Imperial Biomedical Research Centre (BRC), LS is supported by The Parasol Foundation Clinical Senior Lecturer scheme. Additional LC-MS/MS glycomic data were acquired at the Academia Sinica Common Mass Spectrometry Facilities for Proteomics and Protein Modification Analysis, supported by grant AS-CFII-108-107. KHK was supported by an Academia Sinica Investigator Award grant AS-IA-105-L02.

**Author contributions**

S.M.H., A.D., L.S., P.R.B. and D.A.M. conceptualised the study and developed the experimental design. Experiments were performed by G.W., P.G., B.G.M., C.T.H and K.H.K. Data processing, analysis, and interpretation was performed by G.W., P.G., B.G.M., C.T.H and K.H.K., S.K., S.M.H., A.D., L.S., P.R.B. and D.A.M. G.W. and P.G. prepared all figures and tables. G.W., P.G., A.D. and S.M.H. wrote the first draft of the manuscript. All authors critically reviewed, read and approved the final manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20608-7.

**Correspondence** and requests for materials should be addressed to A.D. or S.M.H.

**Reprints and permissions information** is available at www.nature.com/reprints.

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

© The Author(s) 2022