scientific reports

Check for updates

OPEN

Poly-sialylated glycan of cervicovaginal fluid can be a potential marker of preterm birth

Yoon-Young Go[1,2], Gun Wook Park[4], Young Min Hur[1], Young-Ah You[1], Gain Lee[1,2], Rin Chae[1,3], Soo-Min Kim[1,2], Sunwha Park[1] & Young Ju Kim[1,2,3]✉

Preterm birth is a global health issue associated with neonatal death and morbidity. However, current methods of predicting preterm birth are insufficient to accurately screen for risk. This study aimed to assess the potential of site-specific N-glycosylation of cervicovaginal fluid (CVF) proteins as predictive biomarkers of preterm birth in a case-control study. Statistical analysis used Student's t-tests, ROC curve and logistic regression adjusted age and BMI. Using N-glycoproteomic analysis of the CVF, we identified 862 N-glycoproteins in CVF samples form 20 pregnancies and 6595 N-linked glycopeptides used a false discovery rate of less than 1%. Of 173 upregulated glycan in preterm group, we found low levels of fucosylation and high levels of sialylation in preterm birth ($p < 0.05$). Then we found that three poly-sialylated glycans had a high predictive value (AUC = 0.802, $p < 0.017$), which were expressed in all samples. In addition, the glycan model with clinical markers performed better. The results indicate that poly-sialylated glycans in CVF have potential value as novel clinical markers for predicting preterm birth during pregnancy. This study suggests strategies for developing new predictive biomarkers using cervicovaginal glycans to detect preterm birth in advance.

Keywords Cervicovaginal fluid, N-glycoproteomic, Preterm birth, Poly-sialylated glycan, Community state type

Preterm birth is defined as the delivery of a fetus before 37 weeks of pregnancy[1,2]. The causes of preterm birth vary widely; however, approximately 70% occur spontaneously as a result of premature labor and ruptured membranes due to stress, inflammation, infection, or an overactive immune response, and 30% occur in mothers with clinical factors such as preeclampsia, hypertension, or gestational diabetes[3]. Globally, preterm birth rates are increasing every year; with 11% of all births thought to be preterm, an estimated 15 million babies are born preterm worldwide each year[2,4,5]. Preterm birth is a leading cause of neonatal death and serious morbidity, such as neurodevelopmental disorders, learning disabilities, visual impairment, and increased lifetime risk of infectious diseases, and is considered the worst outcome in terms of survival and long-term quality of life for newborns[5]. Despite numerous studies on the underlying pathophysiology of preterm birth, there are currently no reliable diagnostic biomarkers to address preterm birth.

Cervicovaginal fluid (CVF) has been known to be an optimal source to find biomarkers for preterm birth[6]. CVF can be easily obtained from the vaginal wall using noninvasive procedures and is a potential mixture of physiological conditions, immunity, and bacterial communities that affect human pregnancy and parturition[7]. Several proteomic studies of the CVF have revealed a correlation between vaginal dysbiosis, inflammation, and adverse pregnancy outcomes[8–10]. A recent study showed that functionally mixed proteins in human CVF are all highly glycosylated by diverse glycan structures and that glycan expression of CVF glycoproteins plays an important role in host-pathogen recognition, immune response, cellular signaling, and pregnancy status[11].

The occurrence of preterm labor is associated with inflammatory activation from the vaginal–cervix to the maternal-fetal interface, and is generally thought to be due to an ascending infection with pathogenic microorganisms[12–14]. Several previous meta-analyses have consistently demonstrated a correlation between preterm birth and the vaginal microbiome (defined by community state types, CSTs), and found that the CVF of preterm women has a highly diverse vaginal microbiome within a low Lactobacillus CST population[15–17].

[1]Department of Obstetrics and Gynecology, College of Medicine and Ewha Medical Institute, Ewha Womans University, 1071, Anyangcheon-ro, Yangcheon-gu, Seoul 07985, Republic of Korea. [2]Graduate Program in System Health Science and Engineering, Ewha Womans University, Seoul 07985, Republic of Korea. [3]Graduate Program in Division of Artificial Intelligence and Software/Artificial Intelligence Convergence, Ewha Womans University, Seoul 07985, Republic of Korea. [4]Bioinformatics Lab, CellKey Inc, 67 Seocho-daero, Seocho-gu, Seoul, Republic of Korea. ✉email: kkyj@ewha.ac.kr

However, this is contradicted by the clinical perspective that preterm birth occurs in the lactobacillus-rich CST I and conversely, full-term birth occurs in the lactobacillus-poor CST IV[18,19]. Wu et al. analyzed the glycan epitopes involved in the vaginal microbiome-host inflammatory response in CVF and found a pattern of increased fucosylation and sialylation in the preterm birth group[20]. We hypothesized that the glycopeptides of the CVF in pregnant women might differ across CST populations, potentially allowing the identification of specific glycan indicators that can distinguish between term and preterm birth.

In this study, we aimed to assess the potential of site-specific N-glycosylation of CVF proteins as a predictive biomarker of preterm birth. By analyzing N-glycosylation in CVF across different CSTs (CST I, III, IV-A, and IV-B), we identified glycan patterns that could serve as reliable indicators of preterm birth risk. CVF glycan profiling data from each CST population could provide an interpretation of clinical questions and a better understanding to identify predictive biomarkers for preterm labor.

Results

N-glycoproteomic profiling of CVF in pregnant women with term and preterm birth

Twenty CVF samples (full-term birth, $n = 8$; preterm birth, $n = 12$) were used for N-linked glycan analysis (Fig. 1A). The microbial community status in the vagina of each sample was previously characterized according to community state type (CST), and N-glycan profiling was performed in four CSTs groups (CST I, CST III, CST IV-A, and CST IV-B) in this study. Three of the donors in each term and preterm group were in CST I and CST III, which were communities dominated by *Lactobacillus* (CST I: *L. crispatus* and CST III: *L. iners*, respectively). The CST IV consist of anaerobic bacteria, with CST IV-A (- *Gardnerella*) and CST IV-B (+ *Gardnerella*) categorized by the presence or absence of *Gardnerella*. It was not possible to match the CST IV samples between the term and preterm groups because of the limited number of CST IV donors in the term birth group (Supplementary Table S1). Baseline characteristics of all samples were indicated in Table 1. We identified 862 N-glycoproteins, including 203 unique to term samples, 581 shared between term and preterm samples, and 78 unique to preterm samples. A total of 6595 N-linked glycopeptides were quantified with a false discovery rate (FDR) of less than 1% in triplicate. Among these, 3548 glycopeptides were unique to term samples, 1472 were shared between term and preterm samples, and 1575 were unique to preterm samples (Fig. 1B). Of the quantified glycopeptides, 279 were found to be significantly differentially expressed between term and preterm groups (fold change > 2, $p < 0.05$), including 173 upregulated and 106 downregulated peptides. The analysis of differentially expressed N-glycopeptides identified 173 and 106 N-glycopeptides as up- and downregulated in preterm birth group, respectively (fold change > 2, $p < 0.05$) (Fig. 1C). Heatmap analysis showed differentially expressed N-glycopeptides in the four CST groups in term and preterm samples (Supplemental Fig. 1).

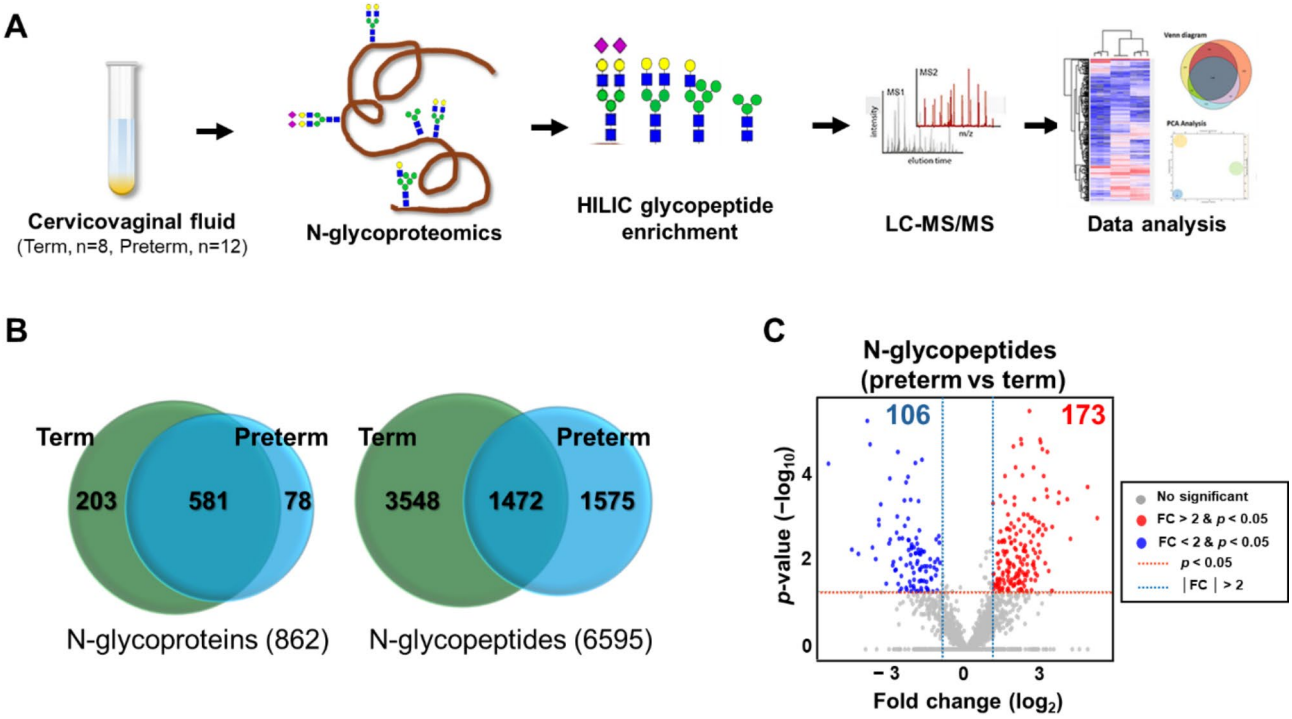

Fig. 1. Overview of analysis of N-glycans isolated from CVF of pregnant women with full-term and preterm birth. (**a**) The schematic workflow for N-linked glycan profiling by hydrophilic interaction chromatography (HILIC). (**b**) Venn diagram presenting identified N-glycoproteins (left) and N-glycopeptides (right) in term and preterm samples. (**c**) Volcano plot shows significantly altered N-glycopeptides (> 2-fold changes, p-value < 0.05).

Characteristics	PTB ($n = 12$)	TB ($n = 8$)	P-value
Maternal age (years)	31.9 ± 3.2	32.6 ± 2.7	0.257
Pre-pregnancy BMI (kg/m²)	21.8 ± 2.8	20.2 ± 1.3	0.240
Nulliparity (n, %)	5 (41.7)	4 (50.0)	0.714
GAS (weeks)	29.0 ± 4.5	30.0 ± 5.2	0.352
GAB (weeks)	31.6 ± 4.1	38.7 ± 1.0	< 0.001*
Birth weight (g)	1913.9 ± 865.0	2998.8 ± 292.7	0.070
Apgar score (1 min)	6.3 ± 2.6	7.5 ± 2.8	0.512
Apgar score (5 min)	8.1 ± 1.7	8.6 ± 1.5	0.902

Table 1. Clinical characteristics of subjects ($n = 20$). PTB, preterm birth; TB, term birth; BMI, body mass index; GAS, gestational age at sampling; GAB, gestational age at birth.

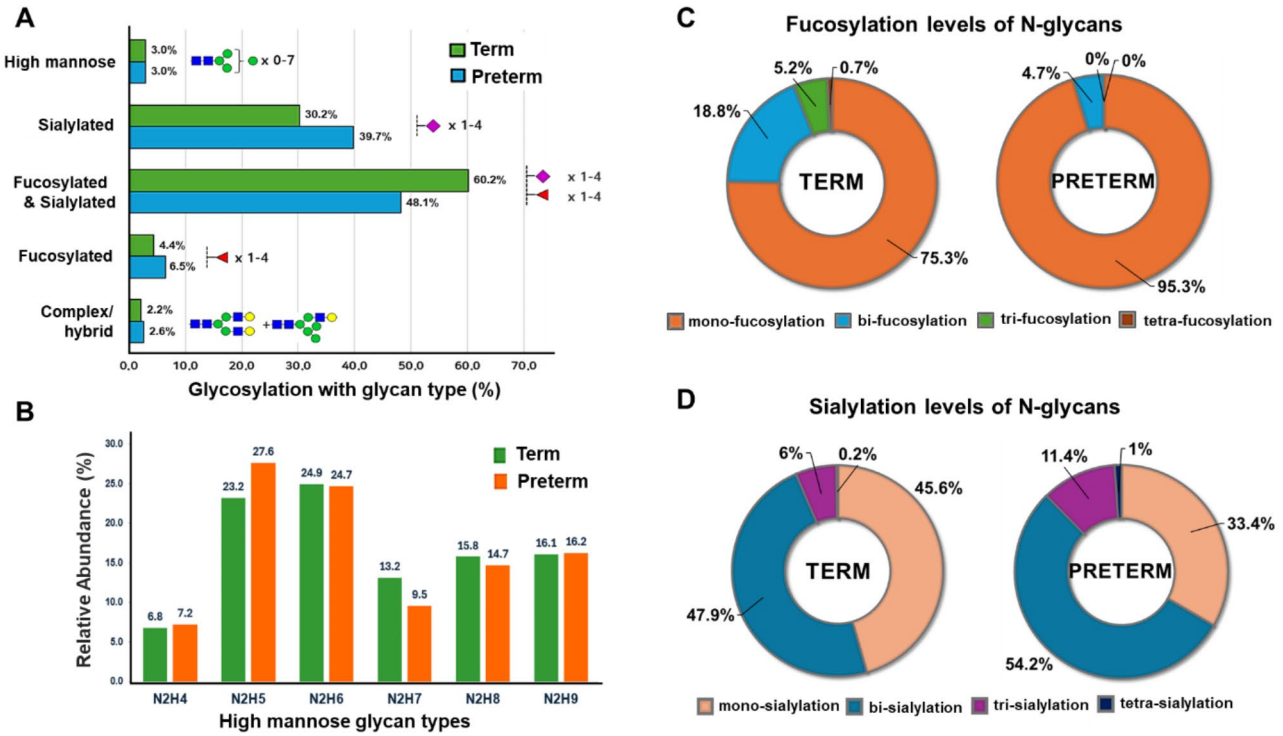

Fig. 2. Variations of the N-glycosylation between term and preterm samples. (**a**) The percentage of five N-glycan types analyze in differentially expressed glycopeptides (p-value < 0.05) between term and preterm groups. (**b**) The relative abundance of high mannose glycan types in term and preterm groups. The proportion of N-glycans (mono-, bi-, tri-, and tetra-) with (**c**) fucosylation and (**d**) sialylation in term and preterm groups.

Differences in the N-glycan between term and preterm birth samples

Next, we explored the comparative N-glycosylation composition of monosaccharides in term and preterm birth samples. N-Glycan structures were categorized into five groups according to their glycoforms: high-mannose, complex/hybrid, fucosylated, sialylated, fucosylated, and sialylated glycans (Fig. 2A). Glycans with sialylation only and those with both sialylation and fucosylation accounted for more than 80% of the N-glycosites in both the term and preterm groups. The percentage of glycosylation with fucosylated only and sialylated only glycans was relatively higher in the preterm birth group than in the term birth group, whereas high-mannose and complex/hybrid glycans were similar between the two groups in this study. The relative occurrence of high-mannose glycans for the same N-glycosites showed little difference between term and preterm birth samples (Fig. 2B). We found that the proportions of glycans with poly-fucosylation and poly-sialylation showed distinct differences between the term and preterm birth groups. Statistical analysis using Student's t-tests revealed significant differences in the mean proportions of these glycan types, with $p < 0.05$, indicating a clear variation between the groups. Compared with the term birth group, glycans with a higher ratio of poly-sialylation and a lower ratio of poly-fucosylation were identified in the preterm birth group (Fig. 2C and D).

Fig. 3. Schematic illustration of specific N-glycosylation in the CVF of the preterm group.

N-glycopeptides (C/H-S glycan)	AUC	*P*-value	SENS	SPEC	PPV	NPV	Accuracy
CTTNBP2 NL EENRTK + HexNAc(5)Hex(6)NeuAc(3)	0.843	0.008	75%	87.5%	90%	70%	80%
ORM2 NEEYNK + HexNAc(5)Hex(6)NeuAc(3)	0.802	0.007	75%	87.5%	90%	70%	80%
SERPINA1 YLGNATAIFFLPDEGK + HexNAc(4)Hex(5)NeuAc(2)	0.875	0.002	75%	87.5%	90%	70%	80%

Table 2. Efficacy of poly-sialylated glycan candidates in predicting preterm birth. AUC; area under the curve, SENS; sensitivity, SPEC; specificity, PPV; positive predictive value, NPV; negative predictive value. CTTNBP2 NL; CTTNBP2 N-terminal like, ORM2; Orosomucoid, Alpha- 1-acid glycoprotein 2, SERPINA1; serine protease inhibitor alpha- 1 antitrypsin.

Characterization of N-glycan in CVF based on microbial composition

We further analyzed the differences in CVF N-glycosylation between term and preterm birth groups according to the four CST statuses. The 65 differentially expressed N-glycopeptides (up: 33, down: 32) were identified in the preterm CST I group compared with the term CST I group (Supplementary Fig. 2A). Comparative analysis of N-glycan structures revealed that the majority of N-glycans were sialylated only, while a smaller proportion contained both sialylation and fucosylation in the term and preterm CST I group. Statistical analysis using Student's t-tests showed significant differences in the mean proportions of these glycan types between term and preterm groups, with $p < 0.05$. The percentages of high-mannose, sialylated, fucosylated, and complex/hybrid glycan types were higher in the preterm CST I samples than in the term CST I samples (Supplementary Fig. 2B). Next, the proportion of poly-fucosylated and -sialylated glycans was compared between the term and preterm CST I groups. The preterm CST I samples exhibited significantly higher levels of mono-fucosylation (95.6%) and poly-sialylation (bi-, tri-, and tetra-sialylations; term CST I: 50.3%, 4.4%, and 0%; preterm CST I: 63.2%, 11.5%, and 1%, respectively). Statistical analysis was conducted using two-sample t-tests to evaluate the differences in glycan proportions between term and preterm groups within CST I, and the results showed significant differences ($p < 0.05$). These findings highlight distinct glycosylation differences between term and preterm CST I group (Supplementary Fig. 2C and D). Along with CST I, CST III showed similar results with a *Lactobacillus* dominant group. The analysis of differentially expressed N-glycans (up: 67, down: 11) in preterm CST III compared to term CST III groups showed that the CVF N-glycan of CST III was composed of similar glycosylation types and proportions as CST I (Supplementary Fig. 3A-D). We also investigated the CVF N-glycosylation of CST IV group, which dominantly consist of anaerobic microbiome and *Gardnerella*. Although the sample size is not adequate to detect a difference between the two study groups, the trend of multiple fucosylation and sialylation levels in term and preterm CST I groups consistent with the CST IV-A and B results (Supplementary Fig. 4).

Validation of specific poly-sialylated glycans for preterm birth prediction

This study analyzed the N-linked glycosylation of human CVF and found that CVF from the preterm group had lower levels of fucosylation and higher levels of sialylation than those from the full-term group, regardless of CST (Fig. 3). To further investigate potential candidates as predictive biomarkers of preterm birth from N-glycosylation analysis, we screened for poly-sialylated glycans that were expressed in all samples among the differentially expressed N-glycopeptides in preterm samples compared to term samples (Supplementary Fig. 5). As a result, we found three specific poly-sialylated glycopeptides that were highly expressed in preterm samples including CTTNBP2 N-terminal like (CTTNBP2 NL) : (EENRTK + HexNAc(5)Hex(6)NeuAc(3)), Orosomucoid, Alpha- 1-acid glycoprotein 2 (ORM2) : (NEEYNK + HexNAc(5)Hex(6)NeuAc(3)), Serine protease inhibitor alpha- 1 antitrypsin (SERPINA1) : (YLGNATAIFFLPDEGK + HexNAc(4)Hex(5)NeuAc(2)). Accuracy in predicting preterm birth was demonstrated by analyzing the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, with all three poly-sialylated glycans showing significant accuracy (AUC > 0.8, $P < 0.017$) (Table 2; Fig. 4a). Among them, NEEYNK + HexNAc(5) Hex(6) NeuAc(3) of ORM2 and YLGNATAIFFLPDEGK + HexNAc(4) Hex(5) NeuAc(2) of SERPINA1 were commonly determined in preterm sample in both CST I and CST IV-B groups with higher peak intensity than those in term sample as multiple sialylated glycans (Supplementary Fig. 6, Fig. 4b). Next, we performed ROC curve analysis to determine their individual and combined predictive performance of signatures. Although the CTTNBP2 NL (EENRTK + HexNAc(5)Hex(6)NeuAc(3)) glycan was not identified in CST I, combining it with the commonly detected SERPINA1 (YLGNATAIFFLPDEGK + HexNAc(4)Hex(5)NeuAc(2)) glycan improved the efficacy of preterm

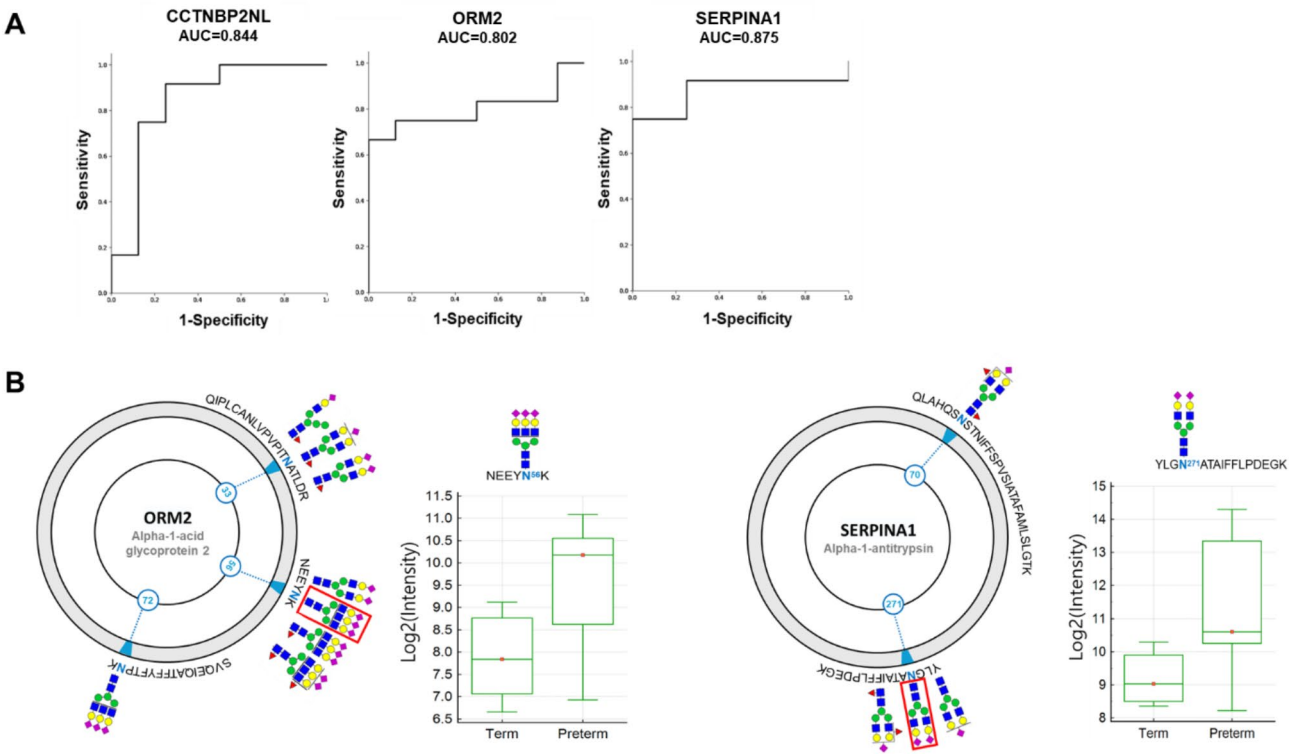

Fig. 4. Predictive performance of specific poly-sialylated glycan candidates for preterm birth prediction. (**a**) Receiver operating characteristics (ROC) curve analysis for specific poly-sialylated glycan candidates (**b**) The representative images of glycopeptides candidates for preterm birth prediction derived from orosomucoid (*ORM2*) and serine protease inhibitor alpha- 1 antitrypsin (*SERPINA1*) encoding genes, respectively. The relative peak intensity of glycopeptides candidates was compared in term and preterm birth groups.

Combinatorial prediction models	AUC	*P*-value	SENS (%)	SPEC (%)	PPV (%)	NPV (%)	Accuracy (%)
CTTNBP2 NL and SERPINA1	0.958	0.001**	83.3	75.0	83.3	75.0	80.0
CTTNBP2 NL, SERPINA1 and Age	0.906	0.003*	84.6	85.7	91.7	75.0	85.0
CTTNBP2 NL, SERPINA1 and BMI	0.958	0.001**	90.9	77.8	83.3	87.5	85.0

Table 3. Predictive validation of a combined model of poly-sialylated glycans and clinical markers using multiple logistic regression. AUC, area under the curve; SENS, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value; BMI, body mass index; WBC, white blood cell. Statistical significance was presented as $P < 0.05$* and $P < 0.001$**.

birth prediction compared to the single models (Table 3; Fig. 5). In addition, clinical markers, such as age and BMI, increased the accuracy, sensitivity, and specificity of the two-glycan combination model.

Discussion

One of the main causes of preterm birth is an inflammatory response activated by an intrauterine infection driven by an ascending genital tract infection[21]. For example, bacterial vaginosis (BV) is a vaginal microbiota imbalance that disrupts immune homeostasis within the host, causing an excessive inflammatory response that leads to adverse pregnancy outcomes such as preterm birth[12,22]. The female reproductive tract is covered with epithelial cells coated with carbohydrates (glycans or glycocalyx), which act as a protective barrier against invading microorganisms[23]. CVF is a glycan-rich biofluid in female genital tract, which is functionally important in determining women's reproductive health and pregnancy outcomes, so glycoproteomic studies of CVF have recently been attempted in obstetrics[11]. BV is known to have low viscosity in the CVF because it contains high levels of glycan-degrading enzymes (glycosidases) and mucolytic enzymes[24,25]. Therefore, the glycosylation analysis of CVF can be used to identify specific biomarkers for predicting preterm birth.

In this study, we performed N-glycopeptide analysis, which offers several advantages, including precise identification of glycosylation sites and structural characterization of glycans attached to specific proteins, deepening our understanding of their functional roles. Wu et al.[11] demonstrate that N-glycosylation profiles in the cervicovaginal fluid (CVF) are intricately linked to pregnancy status, microbial composition, and

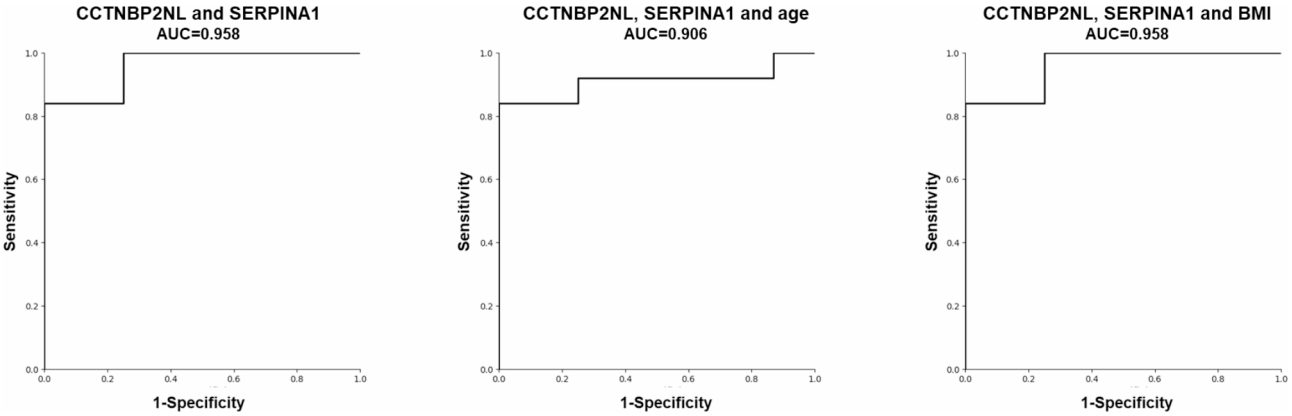

Fig. 5. Predictive validation of a combined model of poly-sialylated glycans and clinical markers using ROC curves.

immune activity, providing insights into the mechanisms of microbial-driven preterm birth. The complexity of glycoproteins, which can exhibit extensive heterogeneity at each N-glycosite, poses a notable challenge for glycoproteomic analysis. Advanced technologies and methodologies have been developed to address these challenges and enable the comprehensive characterization of glycoproteins. We integrated I-GPA[26] and O-GPA[27] to develop a new automated glycoprotein analysis platform, called the SpAC9 Pipeline. This platform has proven efficient, automatic identification, and quantification of large numbers of site-specific N-glycopeptides with low false FDRs, making it a powerful tool for biomarker discovery and glycoproteome analysis[28]. These technologies have opened new avenues for exploring the role of glycoproteome in health and disease, offering detailed insights into glycan structures and their functional implications. The integration of these advanced glycoproteomics tools into obstetric research could lead to the identification of novel biomarkers for preterm birth, ultimately improving clinical outcomes.

Our results showed that the glycosylation composition of CVF differed slightly depending on the CST category, but the conclusion was the same: there were differences in the fucosylation and sialylation levels of CVF in the full-term and preterm groups. The glycan composition of specific N-glycopeptides between term and preterm births showed that the CVF of the preterm birth group had low levels of fucosylation and was highly composed of multiple sialylation levels. Therefore, the occurrence of preterm births in protective CST type 1 with Lactobacillus and full-term births in disruptive CST type 4 B with Gardnerella can be explained by the low fucosylation and high sialylation levels observed in this study. Fucosylation of host glycans is involved in homeostatic mechanisms between host epithelial and immune cells[29]. In particular, the glycoproteomic analysis of human CVF samples showed that low levels of fucosylation were positively correlated with increased levels of pro-inflammatory cytokines, whereas high levels of fucosylation were negatively correlated[11]. Sialylated glycans are commonly associated with infectious events because they promote the binding of bacterial and viral surface proteins to specific sialylation sites in the host[30,31]. These recent studies consistently support our findings and may explain why high levels of sialylation increase interactions with invading pathogens. Subsequently, low levels of fucosylation activate excessive inflammatory responses within the host, leading to preterm birth (Fig. 3).

Wu et al. recently revealed that CVF glycosylation patterns in women at risk of preterm birth show distinct immuno-regulatory epitopes similar to those observed in cancer and viral infections, particularly involving increased sialylation and fucosylation[20]. In contrast to this, our study identified specific sialylated glycans that could predict preterm birth based on CVF N-glycosylation analysis in pregnant women. The accuracy of preterm birth prediction could be improved by applying specific sialylated glycans as biomarkers using machine learning models. Although it did not achieve significance due to the small sample size, CTTNBP2 NL (EENRTK + HexNAc(5) Hex(6) NeuAc(3)), a sialylated glycan selected using a random forest model, used as a promising marker of preterm birth prediction. This could be developed into a novel preterm birth prediction marker as it can be individually measured by multiple reaction monitoring (MRM) analysis[32], which can be used to analyze specific glycosylation changes in the CVF of pregnant women. Although our study had the limitation of a small sample size, it is the first to propose a methodological strategy for the identification and application of candidate biomarkers for the prediction of preterm birth by glycosylation analysis of the CVF. Future research is needed to determine how our work can be applied and developed into an assessment model that can predict preterm births in large populations.

Conclusion

Our results revealed that the polyfucosylation and sialylation levels of CVF were different in the full-term and preterm birth groups. In addition, we identified specific N-linked polysialylated glycans that were elevated in preterm births, and validated their potential as novel biomarkers for predicting preterm birth.

Materials and methods

Study participants and CVF sampling

We previously collected CVF samples in a cohort study to investigate widespread changes in the vaginal microbiome between full-term and preterm births in Korean pregnant women[33]. This study was approved by the Institutional Review Board (IRB) of Ewha Womans University Medical Center (EUMC 2018 - 07-007–010) in accordance with the relevant guidelines and regulations of the Declaration of Helsinki. Briefly, pregnant women with singleton pregnancies were enrolled as study participants, and vaginal swabs were collected during the 21–35 weeks of gestation with written informed consent. The preterm birth group included those who delivered within 37 weeks of gestation and had no medical indications such as placental abruption, short or open cervix, premature rupture of membranes (PPROM), preeclampsia, placenta previa, and intrauterine growth restriction (IUGR).

The vaginal swab was suspended in 2 mL of phosphate buffer saline (PBS) and stored in aliquots at − 80 °C until use.

Sample preparation

The CVF samples were buffer-exchanged and concentrated using a 10 kDa molecular weight cutoff (MWCO) filter (Merck Millipore, Darmstadt, Germany). A bicinchoninic acid (BCA) protein assay (Thermo Fisher Scientific, Waltham, MA, USA) was used to determine the protein concentration. Subsequently, the samples were denatured with 50 mM ammonium bicarbonate at 95 °C and 10 mM dithiothreitol (DTT), followed by alkylation with 20 mM iodoacetamide (IAA) for 45 min at room temperature in the dark. The protein mixtures were digested with trypsin (Promega, Madison, WI, USA) in a 1:50 (enzyme/protein, w/w) ratio for 16 h at 37 °C. N-Glycopeptides were enriched using Hydrophilic Interaction Liquid Chromatography (HILIC) method. HILIC separates molecules based on their hydrophilicity, making it particularly effective for enriching glycopeptides that are highly polar owing to their glycan moieties. Finally, enriched glycopeptides were dried using a SpeedVac concentrator.

Liquid chromatography-mass spectrometry

Samples were analyzed using a Thermo Scientific™ Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer coupled with a Vanquish Neo UHPLC (Thermo Fisher Scientific). All samples were dissolved in loading buffer (0.1% FA in LC-MS grade water), and 3 µL of samples were loaded into a PepMap™ Neo Trap Cartridge (5 µm × 300 µm × 5 mm). The trapped peptides were separated on an EASY-Spray PepMap Neo C18 column (2 µm × 75 µm × 500 mm) at 0.3 µL/min. Mobile phases A and B contained 0.1% formic acid in water and acetonitrile, respectively. The LC gradient began with 2% mobile phase B for 5 min, increased from 2 to 25% B for 85 min, from 25 to 40% B for 10 min, from 40 to 95% B for 5 min, and then held at 95% B for 5 min. The column was then re-equilibrated with 2% B for 10 min. As the peptides were eluted, full scans were acquired in Orbitrap at a resolution of 60,000 with a scan range of 400–2500. MS2 scans were acquired with 33% normalized collision energy (NCE) in high-energy collisional dissociation (HCD) mode at a resolution of 15,000.

Glycopeptide data analysis

The SpAC9 pipeline[27], CellKey's proprietary glycopeptide data analytics solution, was used to automatically identify and quantify site-specific N-glycopeptides. The raw mass spectrometer file from the Orbitrap was converted to an MS (. ms1) and MS/MS (. ms2) files using the freeware program Converter (The Scripps Research Institute, La Jolla, CA)[34]. N-glycopeptides will be analyzed by high resolution mass spectrometry with HCD fragmentation followed by SpAC9, with following search parameters: glycopeptide tolerance = 20 ppm; Fragment tolerance = 0.02 Da (HCD); Missed cleavages = 0, Modification; Carbamidomethyl cysteine (fixed) and Methionine oxidation (variable). A reverse nonsense version of the original database was generated and used to determine the FDR, which was set to 1% for glycopeptide-spectrum matches (GSMs).

Statistical analysis

The basic characteristics of the study groups were analyzed using Student's t-test. The diagnostic accuracy of the glycopeptide candidates for preterm birth was evaluated using AUC and ROC curves, including accuracy, sensitivity, and specificity. Statistical significance was evaluated at $P < 0.05$, or $P < 0.001$**. The Statistical Package for Social Sciences (SPSS, Version 2.0, Chicago, IL, USA) was used for statistical analysis.

Data availability

All data from this study are included in published articles and supplementary information files.

Received: 9 September 2024; Accepted: 31 March 2025
Published online: 03 April 2025

References

1. Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet* **371**, 75–84. https://doi.org/10.1016/S0140-6736(08)60074-4 (2008).
2. Vogel, J. P. et al. The global epidemiology of preterm birth. *Best Pract. Res. Clin. Obstet. Gynecol.* **52**, 3–12. https://doi.org/10.1016/j.bpobgyn.2018.04.003 (2018).
3. Beernink, R. H. J., Schuitemaker, J. H. N., Zwertbroek, E. F., Scherjon, S. A. & Cremers, T. Early pregnancy biomarker discovery study for spontaneous preterm birth. *Placenta* **139**, 112–119. https://doi.org/10.1016/j.placenta.2023.06.011 (2023).
4. Chawanpaiboon, S. et al. Global, regional, and National estimates of levels of preterm birth in 2014: A systematic review and modelling analysis. *Lancet Glob. Health* **7**, e37–e46. https://doi.org/10.1016/s2214-109x(18)30451-0 (2019).

5. Arman, B. M., Binder, N. K., de Alwis, N., Kaitu'u-Lino, T. J. & Hannan, N. J. Repurposing existing drugs as a therapeutic approach for the prevention of preterm birth. *Reproduction* **165**, R9–r23. https://doi.org/10.1530/rep-22-0226 (2023).
6. Cottingham, K. The cervical-vaginal fluid proteome and possible biomarkers of preterm birth. *J. Proteome Res.* **6**, 1241–1242. https://doi.org/10.1021/pr070737o (2007).
7. Dasari, S. et al. Comprehensive proteomic analysis of human cervical-vaginal fluid. *J. Proteome Res.* **6**, 1258–1268. https://doi.org/10.1021/pr0605419 (2007).
8. Kindinger, L. M. et al. Relationship between vaginal microbial dysbiosis, inflammation, and pregnancy outcomes in cervical cerclage. *Sci. Transl. Med.* **8**, 350ra102. https://doi.org/10.1126/scitranslmed.aag1026 (2016).
9. Moncla, B. J., Chappell, C. A., Debo, B. M. & Meyn, L. A. The effects of hormones and vaginal microflora on the glycome of the female genital tract: Cervical-vaginal fluid. *PLoS One.* **11**, e0158687. https://doi.org/10.1371/journal.pone.0158687 (2016).
10. Rizzo, G. et al. Interleukin-6 concentrations in cervical secretions identify microbial invasion of the amniotic cavity in patients with preterm labor and intact membranes. *Am. J. Obstet. Gynecol.* **175**, 812–817. https://doi.org/10.1016/s0002-9378(96)80004-4 (1996).
11. Wu, G. et al. N-glycosylation of cervicovaginal fluid reflects microbial community, immune activity, and pregnancy status. *Sci. Rep.* **12**, 16948. https://doi.org/10.1038/s41598-022-20608-7 (2022).
12. Bayar, E., Bennett, P. R., Chan, D., Sykes, L. & MacIntyre, D. A. The pregnancy microbiome and preterm birth. *Semin. Immunopathol.* **42**, 487–499. https://doi.org/10.1007/s00281-020-00817-w (2020).
13. Romero, R. et al. The role of inflammation and infection in preterm birth. *Semin. Reprod. Med.* **25**, 21–39. https://doi.org/10.1055/s-2006-956773 (2007).
14. Abrahams, V. M. et al. Bacterial modulation of human fetal membrane toll-like receptor expression. *Am. J. Reprod. Immunol.* **69**, 33–40. https://doi.org/10.1111/aji.12016 (2013).
15. Gudnadottir, U. et al. The vaginal microbiome and the risk of preterm birth: A systematic review and network meta-analysis. *Sci. Rep.* **12**, 7926. https://doi.org/10.1038/s41598-022-12007-9 (2022).
16. Fettweis, J. M. et al. The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021. https://doi.org/10.1038/s41591-019-0450-2 (2019).
17. Kosti, I., Lyalina, S., Pollard, K. S., Butte, A. J. & Sirota, M. Meta-analysis of vaginal microbiome data provides new insights into preterm birth. *Front. Microbiol.* **11**, 476. https://doi.org/10.3389/fmicb.2020.00476 (2020).
18. Stafford, G. P. et al. Spontaneous preterm birth is associated with differential expression of vaginal metabolites by Lactobacilli-dominated microflora. *Front. Physiol.* **8**, 615. https://doi.org/10.3389/fphys.2017.00615 (2017).
19. Freitas, A. C., Bocking, A., Hill, J. E. & Money, D. M. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. *Microbiome* **6**, 117. https://doi.org/10.1186/s40168-018-0502-8 (2018).
20. Wu, G. et al. Glycomics of cervicovaginal fluid from women at risk of preterm birth reveals immuno–regulatory epitopes that are hallmarks of cancer and viral glycosylation. *Sci. Rep.* **14**, 20813. https://doi.org/10.1038/s41598-024-71950-x (2024).
21. Gilman-Sachs, A. et al. Inflammation induced preterm labor and birth. *J. Reprod. Immunol.* **129**, 53–58. https://doi.org/10.1016/j.jri.2018.06.029 (2018).
22. Reiter, S. & Kellogg Spadt, S. Bacterial vaginosis: A primer for clinicians. *Postgrad. Med.* **131**, 8–18. https://doi.org/10.1080/00325481.2019.1546534 (2019).
23. Quayle, A. J. The innate and early immune response to pathogen challenge in the female genital tract and the pivotal role of epithelial cells. *J. Reprod. Immunol.* **57**, 61–79. https://doi.org/10.1016/s0165-0378(02)00019-0 (2002).
24. Moncla, B. J. et al. Impact of bacterial vaginosis, as assessed by Nugent criteria and hormonal status on glycosidases and lectin binding in cervicovaginal lavage samples. *PLoS One* **10**, e0127091. https://doi.org/10.1371/journal.pone.0127091 (2015).
25. Howe, L. et al. Mucinase and sialidase activity of the vaginal microflora: Implications for the pathogenesis of preterm labour. *Int. J. STD AIDS* **10**, 442–447. https://doi.org/10.1258/0956462991914438 (1999).
26. Park, G. W. et al. Integrated glycoproteome analyzer (I-GPA) for automated identification and quantitation of site-specific N-glycosylation. *Sci. Rep.* **6**, 21175. https://doi.org/10.1038/srep21175 (2016).
27. Park, G. W. et al. Classification of mucin-type O-glycopeptides using higher-energy collisional dissociation in mass spectrometry. *Anal. Chem.* **92**, 9772–9781. https://doi.org/10.1021/acs.analchem.0c01218 (2020).
28. Kawahara, R. et al. Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nat. Methods* **18**, 1304–1316. https://doi.org/10.1038/s41592-021-01309-x (2021).
29. Kononova, S., Litvinova, E., Vakhitov, T., Skalinskaya, M. & Sitkin, S. Acceptive immunity: The role of fucosylated glycans in human host-microbiome interactions. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms22083854 (2021).
30. Dedola, S. et al. Sialic acids in infection and their potential use in detection and protection against pathogens. *RSC Chem. Biol.* **5**, 167–188. https://doi.org/10.1039/d3cb00155e (2024).
31. Sakai, T., Nishimura, S. I., Naito, T. & Saito, M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Sci. Rep.* **7**, 45043. https://doi.org/10.1038/srep45043 (2017).
32. Delafield, D. G. & Li, L. Recent advances in analytical approaches for glycan and glycopeptide quantitation. *Mol. Cell. Proteom.* **20**, 100054. https://doi.org/10.1074/mcp.R120.002095 (2021).
33. You, Y. A. et al. Transition in vaginal Lactobacillus species during pregnancy and prediction of preterm birth in Korean women. *Sci. Rep.* **12**, 22303. https://doi.org/10.1038/s41598-022-26058-5 (2022).
34. He, L., Diedrich, J., Chu, Y. Y. & Yates, J. R. 3 Extracting accurate precursor information for tandem mass spectra by rawconverter. *Anal. Chem.* **87**, 11361–11367. https://doi.org/10.1021/acs.analchem.5b02721 (2015).

Acknowledgements

We thank the Korea Health Industry Development Institute (RS- 2023 - 00262969) of the UroGyn Efficacy Evaluation Center, Institute of Convergence Medicine, Ewha Womans University, Mokdong Hospital, Seoul, Korea.

Author contributions

YYG, YMH, and YAY designed the study concept and methodology; GWP, YYG, GL, SMK, and RC performed statistical analysis and visualization; YYG, GWP, and SWP interpreted and analyzed the data; YYG and GWP wrote the manuscript; YJK supervised the study.

Funding

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: RS- 2023 - 00266554).

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
The present study was approved by the ethics committees of all participating hospitals where the pregnant women were recruited.

Consent to participate
All study participants provided written informed consent prior to enrolment.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96682-4.

Correspondence and requests for materials should be addressed to Y.J.K.

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.