# Unlocking uterine biology at home: a validated system for DNA, RNA, and microbial analysis from menstrual effluence

***Stephen Gire***[1], *Xitong Li*[1], *Claire Toth*[1], *Meet Doshi*[3], *Sachin Kumar Gupta*[3], *Scott Parker*[2], *Debbie Boles*[2], *Ridhi Tariyal*[1]

1. NextGen Jane, Inc. Oakland, California
2. Labcorp, Research Triangle Park, North Carolina
3. Elucidata, San Francisco, California

Corresponding Authors:

Stephen Gire: stephen@nextgenjane.com

Ridhi Tariyal: ridhi@nextgenjane.com

## ABSTRACT

**Introduction:** Access to accurate, non-invasive diagnostics remains a critical unmet need in women's health. Menstrual effluence, containing endometrial tissue, immune cells, and microbial communities, represents a clinically relevant specimen for genomic and molecular pathology applications, yet has historically been underutilized due to concerns about sample integrity and variability.

**Methods:** We developed and validated a standardized, at-home tampon-based collection system designed to preserve nucleic acids at ambient temperature for clinical-grade analyses. 1,067 tampon samples from 328 participants underwent, RNA sequencing and metatranscriptomic profiling to assess specimen transcript integrity, diagnostic fidelity, and microbial composition over time. 12 patients were exome sequenced using matched menstrual effluence and whole blood to assess assay concordance between sample types.

**Results:** RNA extracted from menstrual effluence maintained stability for up to 14 days without refrigeration, achieving sufficient yield and quality for sequencing in >97% of samples. Variant detection via exome sequencing demonstrated 100% concordance among overlapping single nucleotide variants between menstrual fluid and matched venous blood, confirming clinical equivalency for genetic testing. Transcriptomic analyses revealed cycle-dependent variation in key reproductive and immune markers, while metatranscriptomic profiling identified shifts in microbial communities consistent with known reproductive tract dysbiosis.

**Conclusions:** Standardized at-home collection of menstrual effluence provides a clinically actionable platform that supports remote specimen acquisition without compromising molecular assay fidelity, offering a scalable solution to improve access to carrier screening, reproductive health assessment, and infectious disease monitoring in clinical practice.

## INTRODUCTION

Women's health has been underserved by non-specific diagnostics and limited therapeutic

39 options. Despite high disease burden, conditions like endometriosis, fibroids, and recurrent
40 miscarriage are diagnosed and treated with tools that lack precision, leading to significant delays
41 in care[1]. Broader reliance on non-specific clinical pathways has led to a one-size-fits-all approach
42 to care[2–4]. For instance, gonadotropin-releasing hormone (GnRH) analogs—despite inducing
43 premature menopause and causing severe side effects such as bone loss and mood
44 disturbances[5] – remain a first-line treatment for endometriosis[6] and fibroid-related heavy
45 bleeding[7]. CA125, a non-specific marker of inflammation, is monitored across a wide range of
46 conditions including ovarian cancer[8], heart failure[9], miscarriage risk[10], and endometriosis[11],
47 despite limited diagnostic specificity[12]. Without condition-specific diagnostics, targeted therapies
48 have remained underdeveloped, perpetuating inadequate care across puberty, pregnancy, and
49 menopause.
50
51 At-home biospecimen collection systems offer a scalable way to improve diagnostics and enable
52 precision medicine in women's health. Remote collection of biological samples allows clinical data
53 to be captured outside of brick-and-mortar healthcare settings, improving access and enabling
54 earlier decision-making[13]. For research, these systems offer opportunities to build rich,
55 longitudinal, high-dimensional datasets that are essential for studying reproductive biology at
56 scale. However, few such datasets exist, and those available are poorly harmonized[14], limiting
57 their cross-study comparability and value for biomarker discovery.
58
59 Menstrual effluence is a rich but underutilized biospecimen that closely mirrors the endometrial
60 environment and presents novel opportunities for diagnostic development. Menstrual fluid is
61 expelled naturally and contains a mix of shed endometrium, interstitial fluid, blood, mucosa, and
62 epithelial cells from the cervix and vagina, along with reproductive tract microbiota[15,16]. Organoids
63 derived from menstrual effluence demonstrate similar transcriptomic profiles and hormone
64 responsiveness as those from invasive endometrial biopsies[17]. Importantly, menstrual effluence

65 contains a unique population of immune cells enriched for innate and adaptive subsets involved
66 in tissue remodeling and immune tolerance[18], unlike peripheral blood. The sample's heterogeneity
67 captures a broader snapshot of uterine and vaginal health than traditional biopsies or swabs.
68 Despite its promise, menstrual effluence has historically been viewed as technically difficult to
69 work with. Approximately 30% of cells in menstrual fluid are non-viable[15,16], and tissue aggregates
70 and mucosal debris complicate laboratory workflows. Additionally, cellular lysis releases
71 nucleases that can degrade nucleic acids, compromising downstream molecular assays. Sample
72 composition also varies by collection method[19] – e.g., pad vs. tampon vs. menstrual cup –
73 introducing variability and contributing to skepticism about its reliability as a biospecimen.
74 However, these challenges are surmountable with proper pre-analytical protocols.
75
76 New advances have transformed how menstrual effluence is viewed in diagnostics and research.
77 Recent studies have generated high-resolution datasets from menstrual fluid and uterine tissue
78 that offer insight into reproductive pathophysiology[15,16,20]. Transcriptomic data from menstrual
79 effluence have revealed dynamic biological processes such as epithelial-to-mesenchymal
80 transition relevant to endometriosis progression[21]. Endotoxins detected in menstrual samples
81 have been linked to adverse pregnancy outcomes[22], and menstrual immune profiles have shown
82 potential for diagnosing inflammatory and infectious conditions[23]. These discoveries demonstrate
83 that menstrual effluence is not only clinically relevant but may offer information inaccessible via
84 other sample types.
85
86 We developed a standardized, tampon-based system to enable high-quality menstrual fluid
87 collection at home. To address pre-analytical variability and improve sample stability, we built a
88 self-contained tampon collection kit that also captures metadata such as cycle day and bleeding
89 volume, which serve as important co-variables in transcriptomic analyses. Prior research has
90 shown that aligning sample timing with menstrual phase improves diagnostic accuracy in

91 cardiovascular[24] and mental health applications[25]. The NextGen Jane® (NGJ) platform enables
92 consistent sample collection, better supports hormonal context interpretation, and allows for
93 widespread use in both clinical and research settings.
94
95 We validated our tampon-based sample collection system through large-scale deployment and
96 demonstrated its feasibility and diagnostic utility in three clinically relevant applications through
97 DNA exome sequencing (carrier screening) and unbiased RNA sequencing of human and
98 microbial cell types (molecular phenotyping and microbial surveillance). We have collected 2,025
99 RNA-seq and metatranscriptomic datasets from 1,067 tampons submitted by 328 U.S.
100 participants. All samples were home-collected and shipped at ambient temperature to a
101 centralized lab in Oakland, CA, without loss of data quality – demonstrating the viability of
102 decentralized sampling for molecular diagnostics.
103
104 **MATERIALS AND METHODS**
105 A thorough accounting of materials and methods can be found in Supplemental Methods.docx
106
107 **Patient Recruitment and Metadata Collection**
108 Patients were recruited under IRB-approved protocols (20192619, 20191947, 20233438).
109 Consent was obtained from participants who agreed to tampon-based specimen collection and
110 completed a survey on clinical history, menstrual regimen, and demographics. Data were
111 organized in four tiers: patient-level (race, ethnicity), cycle-level (age, geography, birth control),
112 tampon/kit-level (flow type, cycle day), and sample-level (RNA concentration, library size).
113 Sample accession and collection dates were recorded to assess sample stability over time. Flow
114 rate was calculated from the weight of the tampon before and after collection. Metadata can be
115 found in Supplemental_RNASeq.xlsx. Patient IDs C01-, H01, E01), Kit IDs (NAPS) and Sample
116 IDs (ESGs), are randomized and anonymized ID's that cannot be tied back to patients outside

117 the research team.

118

119 **Sample Collection and Processing**

120 Participants used organic cotton tampons during multiple days of menstruation, wearing them for
121 4 hours. After collection, the tampon was sealed in a collection jar with 20mL of preservation
122 buffer (Norgen Biotek). Samples were sent back to the lab via USPS. Upon receipt, samples were
123 processed, including extrusion, centrifugation, and aliquoting into cryovials for storage at -80°C.

124

125 **Hemoglobin Content Spectrophotometry**

126 The tampon sample was diluted 1:4 with a preservation buffer and analyzed using a UV/Vis
127 spectrophotometer to measure absorbance at 550nm.

128

129 **Nucleic Acid Extraction and Quality Assessment**

130 RNA and DNA were extracted using the Norgen Column-Based RNA Extraction and MagMax
131 mirVana Total RNA Isolation methods. For RNA, DNase treatment was applied, followed by clean-
132 up with RNA XP clean beads. DNA was extracted using the MagMax-96 DNA Multi-Sample Kit.
133 The extracted nucleic acids were quantified using a Qubit 4.0 fluorometer.

134

135 **Exome Sequencing and Carrier Screening**

136 DNA exome sequencing was outsourced to Labcorp for analysis under their established
137 protocols. SNVs and sample metadata can be found in Supplemental_DNASeq.xlsx

138

139 **RNA Sequencing and Analysis**

140 RNA libraries were prepared using the Zymo-Seq RiboFree Total RNA Library Kit, followed by
141 sequencing on an Illumina NextSeq2000. Data processing included adapter trimming, alignment
142 to the hg38 human genome, and gene count normalization using STAR Aligner and

143 FeatureCount. RNA degradation effects were accounted for using the DegNorm program.
144 Sequencing metadata and gene expression values can be found in Supplemental_RNASeq.xlsx.
145 Tampon and menstrual cup samples were compared to publicly available RNA sequencing data
146 on GTeX to calculate Spearman correlations between tissue types. Samples used for this analysis
147 can be found in Supplemental_GTeX.xlsx. Differential expression was performed in XLSTAT.
148

149 **Microbiome Analysis**

150 Metatranscriptomic reads were aligned to microbial databases using Kraken2 to generate raw
151 read counts. For microbiome diversity, species-level and genus-level counts were aggregated to
152 genus level and analyzed using log-transformation and CPM normalization. Sequencing
153 metadata and gene expression values can be found in Supplemental_RNASeq.xlsx
154

155 **Metadata and Statistical Analysis**

156 Metadata, including clinical phenotypes and sample characteristics, were used in ANOVA to
157 assess statistical significance across various groups. Both pvalue and cohen's f for effect size
158 were calculated for all analyses using XLSTAT. Analysis focused on bleeding phenotypes, flow
159 rates, and other collection-related variables.
160

161 **RESULTS**

162 **A tampon-based at-home collection system can reproducibly collect menstrual effluence**
163 **for downstream molecular analyses under real-world mailing and storage conditions.**

164 The NGJ tampon collection system was designed for ease-of-use, integration into daily routines,
165 and standardization to ensure reliable biological preservation. Participants use the provided
166 TOTM-brand organic, low-absorbency tampons, minimizing synthetic additives and heavy metals.
167 Tampons are worn for four hours – reducing toxic shock risk and facilitates sample processing.
168 The cardboard applicator reduces contact with the absorbent core, limiting commensal

169 contamination. A nitrile glove is included for removal and placement into the jar. Once sealed, the
170 cap pulls the string into the jar, preventing leaks, and a shuttle punctures a foil seal to release the
171 preservation buffer, supersaturating the tampon. The sample is bagged, intake questions
172 completed and returned via standard mail. User-tested instructions guide participants through the
173 collection process. A picture of the collection kit and cover of the instructions for use can be found
174 in Figure 1.
175
176 *Demographic Data of Sample Population* – We analyzed menstrual samples collected on cycle
177 days 1–3 with complete phenotypic and technical data (Supplementary Table 1). Inclusion criteria
178 ensured samples were menstrual (not off-cycle vaginal tampons), had complete metadata (flow
179 type, collection time, bleeding phenotype), and passed sequencing quality checks (strandedness,
180 ribosomal rRNA content and uniquely mapped reads). This resulted in 1,135 sequences from 601
181 tampons across 276 participants.
182
183 Table 1 summarizes participant demographics and key flow characteristics. 75% (207
184 participants) identified as Non-Hispanic White, with the remaining 69 participants identifying as
185 Asian, Black, or more than one race. Of 365 cycles, 287 were from individuals not using birth
186 control. Enrolled participants were asked to report the ***flow phenotype*** (overall period heaviness),
187 characterized as heavy, medium or light, corresponding to their subjective experience of their
188 entire menstrual cycle at the time of collection. ***Flow type*** (heaviness on the collection day) was
189 also reported by participants. These measures differ – e.g., a heavy-flow tampon from a light
190 bleeder may differ significantly from one of a heavy bleeder.
191
192 ***Menstrual effluence preserved in stabilization buffer maintains sufficient RNA integrity for***
193 ***sequencing for at least 14 days at ambient temperature.***
194 Home-based sample collection enables large-scale, longitudinal studies across diverse locations.

195 We report nucleic acid yields from our tampon collection system and transcript degradation index
196 (DI), similar to the transcript integrity number (TIN)[26,27]. For sequencing, 11% of samples failed at
197 least once. Among initially failed samples, 79% were successfully resequenced, yielding a final
198 estimated sample failure rate of 2.3%. Approximately 60 million reads are generated per sample
199 (median = 60 million ± 23 million reads), which include both human and microbial transcripts.
200 Nucleic acid extracted from a 600ul aliquot of sample yields enough RNA for downstream
201 sequencing. After lysate release from tampons, an average of 17 mL of total volume
202 (mean = 17.13 ± 2.73) was recovered, providing sufficient material for downstream assays. A
203 500 ng RNA threshold – required for Zymo RiboFree sequencing – was routinely achieved.
204 Although this threshold is reachable via deep-well extraction, optimizing magnetic bead protocols
205 significantly improved nucleic acid recovery (Table 2), enabling use of standard plate formats.
206 Using 600 µL of lysate, 500 ng RNA was obtained in 93% of samples, with a median RNA yield
207 of 1.5 µg.
208
209 We assessed transcript degradation during ambient-temperature storage by measuring the DI of
210 the 25th percentile of genes using DegNorm[26]. RNA degradation causes 3′ coverage bias in RNA-
211 seq data. The DI [26] quantifies this degradation at the gene level, with higher scores reflecting
212 greater 3′ bias. Based on sample collection and processing dates, samples spent an average of
213 6.7 ± 3.52 days in buffer. No significant change in 5′–3′ coverage bias (DI_25) between days 1
214 and 21 (Figure 2a), indicating stability at typical shipping conditions. We further assessed the
215 impact of menstrual flow rate on RNA quality. Flow rate had a strong effect on DI_25: participants
216 with slower flows (<1 grams/hr) had higher degradation, whereas faster flow rates (1–2 grams/hr
217 or >2 grams/hr) correlated with reduced degradation (p=0.002; cohen's f = 0.230 Figure 2b).
218 Extended collection times in slow-flow individuals may exacerbate degradation.
219
220 No seasonal effects on DI_25 were observed (Figure 2b). We did detect significantly lower overall

221 degradation in samples collected in the Midwest region of the United States ($p = 0.023$; cohen's f
222 = 0.127) (Figure 2d). While not statistically significant, samples from the Midwest also show an
223 increase in mean flow rate compared to other regions (Midwest mean= 1.423 g/hr; Other Region
224 mean = 1.149 g/hr; $p = 0.090$; cohen's f = 0.110). This could explain the decrease in DI_25 scores
225 for this region. A total of 11 kits were successfully sequenced at least three times by separate
226 technicians, showing a median coefficient of variation between replicates of 0.149 (supplemental
227 data).
228
229 From 600 µL lysate, we consistently recovered sufficient RNA and DNA for sequencing. RNA
230 remains stable for at least two weeks in buffer at ambient temperatures, showing minimal
231 transcript degradation (see Figure 2a,b) – making this system ideal for at-home collection. With
232 17 mL total lysate per tampon (equivalent volume to extract and sequence the sample 28 times),
233 ample material remains for molecular assays and biobanking. Collectively, we have biobanked
234 2,013 tampon-derived samples.
235
236 **DNA from tampon-collected menstrual effluence yields variant calls that are concordant**
237 **with those obtained from matched blood-based exome sequencing.**
238 At-home, longitudinal collection of menstrual effluence offers accessible prenatal and antenatal
239 screening, especially in gynecological deserts or where remote sampling enhances clinical care[28].
240 To assess concordance of single nucleotide variants (SNVs) between menstrual effluence and
241 venous blood, both sample types were collected from 12 participants and exome sequenced via
242 Labcorp's Inheritest.
243
244 Percentage of overlapping SNVs between blood and menstrual samples was high
245 (mean = 0.983 ± 0.0089), which is similar to other studies showing high concordance of SNVs
246 between different sample types from the same patient[29]. On average, menstrual effluence

247 detected 143 more SNVs than venous blood. Non-overlapping SNVs had lower coverage than
248 overlapping SNVs, likely explaining the difference (median coverage depth for overlapping
249 SNVs = 103 vs. non-overlapping SNVs = 40). Four participants carried pathogenic variants
250 detected in both sample types: GJB2:p.Gly12fs (hearing loss), CFTR:p.Ile507_Phe508delinsIle
251 (cystic fibrosis), CAPN3:p.Arg489Gln (limb-girdle muscular dystrophy), and HEXA:p.Arg499Hi
252 (Tay-Sachs), with allele frequencies ranging from 38% to 62% (Figure 3).
253
254 ***The transcriptomic profile of menstrual effluence represents a unique composite of***
255 ***uterine, vaginal, and immune-derived cells distinct from isolated reproductive tissues.***
256 To assess tissue similarity, we compared RNA expression profiles of menstrual effluence
257 collected via tampon (MB) and menstrual cup (MC) to GTEx data. Reproductive tissues (ovary,
258 vagina, fallopian tubes, cervix, uterus), whole blood, and small intestine (as outgroup) were
259 included. Spearman correlation was used to evaluate global similarity. Differential expression was
260 performed between menstrual effluence and GTEx whole blood samples to determine overlapping
261 genes between menstrual and reproductive tissues. The top 25 overexpressed genes in
262 menstrual effluence show some overlap with reproductive tissues in GTEx (Figure 4b),
263 demonstrating both shared and unique gene programs between the tissues. MB and MC samples
264 were highly correlated (Spearman's r = 0.97), with key differences in keratinized epithelium
265 expression (Supplemental Figure 1a), likely due to vaginal wall cell inclusion in tampon samples[19].
266 Reproductive tissues showed low correlation (r = 0.23–0.32), comparable to small intestine or
267 whole blood (r = 0.45; Figure 4a). This suggests that menstrual effluence shares only a limited
268 transcriptional overlap with reproductive tissue biopsies, similar to the non-reproductive tissues
269 included in the analysis. Differences in cycle phase and sequencing technology of GTEx data
270 may also contribute to this low correlation. The mixture of blood, uterine, and vaginal tissue could
271 represent shared gene programs between multiple tissue types leading to low correlations to any
272 one tissue.

**Gene expression signatures in menstrual effluence vary according to user-reported hormonal use, flow rate, and cycle timing, enabling molecular phenotyping.**

Endometrial gene expression provides valuable insight into menstrual physiology and molecular phenotypes, spanning objective metrics like flow rate (grams/hour) to subjective variables such as self-reported cycle day and flow type. To identify potential molecular correlates, we curated 89 endometrial-enriched genes from the Human Protein Atlas. Of these, 64 were detected in our sequencing data, with 30 genes showing JaneScore-normalized expression ≥2. Notable examples include canonical endometrial markers such as *PAEP*, *ESR1*, and several matrix metalloproteinases (*MMP10*, *MMP11*). The genes not present or at low expression may represent inherent differences between menstrual effluence and endometrium or be due to varying approaches in RNA sequencing.

Flow rate varies by birth control and subjective flow phenotype. Patients using any form of birth control exhibited reduced menstrual flow, as determined from 569 samples with reported data (Figure 5a). When flow rate was combined with spectrophotometric hemoglobin measurements at 550 nm, self-identified heavy bleeders (n = 83) showed significantly darker, more hemoglobin-rich effluence ($p < 0.0001$; cohen's $f=0.352$); Figure 5b). This single-tampon assay presents a scalable alternative to the burdensome alkaline hematin method, which requires collection of all products used over an entire cycle.

Genes enriched in the endometrium correlate with cycle day and flow type. Expression of *GNLY*, a broad-spectrum antimicrobial protein, varied significantly by cycle day ($p < 0.0001$; cohen's $f=0.394$; Figure 5c), indicating immune modulation across menstruation. Furthermore, self-reported flow heaviness (Heavy, Medium, Light, Spotting) was strongly associated with *PAEP* expression in tampon samples ($p < 0.0001$; cohen's $f= 0.469$; Figure 5d).

299

300 ***Menstrual effluence contains transcriptionally active microbial signatures that vary with***
301 ***menstrual day and may include clinically relevant pathogens.***

302 Menstrual samples show a time dependent dysbiosis in microbial transcription with an increase
303 in BV causing bacteria on days 2 and 3 of menstruation. RNA transcriptional activity is distinct
304 from the presence (relative abundance) of microbial DNA. We performed genus-level
305 metatranscriptomic analysis using Kraken2 on RNA reads. We identified high transcriptional
306 activity from several bacterial genera commonly associated with vaginal and uterine microbiota,
307 including *Prevotella*, *Gardnerella*, *Corynebacterium*, *Finegoldia*, and *Anaerococcus* (Figure 6a).
308 These genera are often implicated in microbial dysbiosis and reproductive tract infections.
309 *Lactobacillus* species – typically associated with a healthy reproductive tract – were found to
310 inversely correlate with the abundance of other bacterial taxa. In samples where *Lactobacillus*
311 activity was high, the activity of other bacteria was generally suppressed. However, many samples
312 exhibited low *Lactobacillus* levels (sum of all *Lactobacillus* species below 50% in 746 out of 1,135
313 sequences). alongside elevated expression of potentially pathogenic or opportunistic species.
314 Interestingly, total *Lactobacillus* transcriptional activity decreased significantly on menstrual days
315 2 and 3 (Figure 6b), suggesting a temporal shift in microbial composition during menstruation.
316 These findings present an opportunity to explore and understand typical microbial transitions
317 during a single menstrual cycle and map how vaginal microbiome returns to equilibrium after a
318 menstrual disruption.

319

320 **DISCUSSION**

321 This study presents a standardized, at-home tampon-based system for the collection and
322 stabilization of menstrual effluence, offering a scalable approach for molecular diagnostics and
323 research in women's health. The system preserves RNA and DNA for at least two weeks without
324 refrigeration, supporting applications such as genetic carrier screening, RNA and exome

325 sequencing, and microbial profiling. Its simplicity and suitability for at-home sampling help
326 overcome persistent barriers to diagnostic access, especially in underserved populations.
327 Through the NGJ platform, patients seeking to get pregnant could access their genetic screening
328 results and uterine microbiome insight from a single sample collected at home.
329
330 A key contribution of this work is the ability to maintain nucleic acid integrity in a biologically
331 complex, degradation-prone matrix. Menstrual effluence contains a heterogeneous mixture of
332 blood, endometrial tissue, vaginal secretions, and immune cells, which contribute to enzymatic
333 degradation. By limiting tampon wear to four hours, using chemically inert, low-absorbency cotton
334 tampons, and stabilizing samples in a high-performance buffer, we preserved high-quality RNA
335 and DNA for downstream molecular applications. Higher menstrual flow rates correlated with
336 improved RNA quality, likely due to greater blood content and reduced degradation time.
337 Post-collection experiments further underscored the importance of immediate stabilization, as
338 RNA integrity declined significantly in samples exposed to 37°C for just 30 minutes before
339 preservation (Supplementary Figure 1). These findings provide actionable insights for real-world
340 implementation of mail-in collection systems.
341
342 From a systems biology perspective, our findings highlight menstruation as a dynamic and
343 immunologically active window into reproductive health. Transcriptomic profiles varied across the
344 cycle and by hormonal use, with genes such as *PAEP*, *ESR1*, *GNLY*, and *MMP10* reflecting
345 hormonal withdrawal, immune activation, and tissue remodeling. These changes are difficult to
346 capture via biopsy or swabs, underscoring the value of menstrual effluence for temporal molecular
347 profiling that should be accounted for in research and clinical findings[30–32].
348
349 The high concordance between venous blood and menstrual effluence in exome sequencing
350 supports its use as a non-invasive alternative for carrier screening. Variants linked to conditions

351 such as cystic fibrosis and Tay-Sachs were concordantly detected in both sample types,
352 demonstrating that menstrual effluence can expand access to genomic testing.
353
354 Our system also links subjective menstrual experiences with objective molecular measurements.
355 Variables such as flow heaviness and cycle day correlated with RNA gene expression patterns,
356 suggesting tampon-derived metrics could serve as proxies for subjective reporting or hematin-
357 based methods in heavy menstrual bleeding classification.
358
359 Beyond transcriptomics, metatranscriptomic profiling revealed day-specific shifts in microbial
360 communities. Decreases in *Lactobacillus* and increases in *Gardnerella* and *Prevotella* during
361 menstruation suggest a transient dysbiosis. Coupling host transcriptional responses with
362 microbial activity enables nuanced assessment of conditions like bacterial vaginosis or chronic
363 endometritis, which affect a significant proportion of infertile individuals[33]. Immune activation
364 markers may help determine whether antibiotic treatment is warranted in asymptomatic cases[34,35].
365 Our system allows for this dual measurement in a single, non-invasive specimen. RNA
366 transcriptional profiles of the menstrual microbial community expand the potential of menstrual
367 effluence as a non-invasive monitoring tool for chronic infections, sexually transmitted infections,
368 or dysbiosis-associated conditions like pelvic inflammatory disease.
369
370 The reproducibility and standardization of this collection method also make it well-suited for
371 research at scale. By capturing molecular profiles tied to menstrual phenotypes, this platform can
372 enable the development of classifiers for endometriosis, fibroids, and other gynecologic
373 conditions. Future work will endeavor to utilize this system to better understand the molecular
374 characteristics of menstrual effluence and how variation in cycle data within patients and between
375 patients may influence normal or abnormal menstruation.
376

## CONCLUSION

In conclusion, the development and validation of a standardized tampon-based system for menstrual effluence collection marks a significant advancement in women's health diagnostics. Our findings demonstrate that menstrual effluence, a sample previously considered suboptimal for molecular analysis, can be preserved effectively for RNA and DNA extraction, enabling genetic screening, molecular phenotyping, and pathogen surveillance. This system offers unique opportunities for at-home, longitudinal monitoring of reproductive health, facilitating both clinical care and research applications.

The ability to link menstrual biomarkers with clinical outcomes, such as genetic carrier status or microbial dysbiosis, provides new avenues for early detection and personalized treatment strategies. By providing reliable, reproducible molecular data across diverse menstrual phases and bleeding phenotypes, this approach enhances the potential for diagnostic tools that address the unique health needs of women. Future work will focus on refining the system's integration into clinical practice, developing molecular classifiers for reproductive diseases like endometriosis, and expanding the applications of menstrual effluence in broader health research.

**Conflicts of Interest**

Ridhi Tariyal, Stephen Gire, Xitong Li, and Claire Toth are or were employed by NextGen Jane, Inc. which manufactures the tampon collection kit. The other authors did not report any potential conflicts of interest.

**Funding Sources**

This work was funded in part through an SBIR FastTrack grant from NICHD. Unique Federal Award Identification Number R44HD103159.

**Acknowledgements**

We would like to thank Stephen Palmer, Ruth Kulicke, Matt Stremlau, Ryan Phan, and Pardis Sabeti for their helpful edits. We would like to thank the LabCorp team that helped coordinate exome sequencing efforts: Marcia Eisenberg, Christina Smith, and Megann Vaughn. We would like to thank Abhishek Jha and Yogesh Lakhotia for bioinformatic support. Margaret Eisen ,Julia Carr and Aparna Kola coordinated all clinical collection efforts for this study. The analyses in this paper were partly funded through an SBIR Fast Track grant with the National Institute of Child Health and Disease *5R44HD103159.*

## REFERENCES


1. Cromeens MG, Carey ET, Robinson WR, Knafl K, Thoyre S. Timing, delays and pathways to diagnosis of endometriosis: a scoping review protocol. BMJ Open. BMJ; 2021 Jun 24;11(6):e049390. PMCID: PMC8231065
2. Domar A, Vassena R, Dixon M, Costa M, Vegni E, Collura B, Markert M, Samuelsen C, Guiglotto J, Roitmann E, Boivin J. Barriers and factors associated with significant delays to initial consultation and treatment for infertile patients and partners of infertile patients. Reprod Biomed Online. Elsevier BV; 2021 Dec;43(6):1126–1136. PMID: 34756644
3. Fryer J, Mason-Jones AJ, Woodward A. Understanding diagnostic delay for endometriosis: a scoping review [Internet]. Obstetrics and Gynecology. medRxiv; 2024. Available from: https://www.medrxiv.org/content/10.1101/2024.01.08.24300988v1.full
4. Regensteiner JG, McNeil M, Faubion SS, Bairey-Merz CN, Gulati M, Joffe H, Redberg RF, Rosen SE, Reusch JE, Klein W. Barriers and solutions in women's health research and clinical care: a call to action. Lancet Reg Health Am. Elsevier; 2025 Apr;44:101037. PMCID: PMC11937288
5. Gonadotropin releasing hormone (GnRH) analogues. LiverTox: Clinical and Research Information on Drug-Induced Liver Injury. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; 2012. PMID: 31643199
6. Saleh FL, Taylor HS. Clinical applications of gonadotropin-releasing hormone analogues: a broad impact on reproductive medicine. F S Rep. Elsevier BV; 2023 Jun;4(2 Suppl):83–87. PMCID: PMC10201293
7. Madueke-Laveaux OS, Ciebiera M, Al-Hendy A. GnRH analogs for the treatment of heavy menstrual bleeding associated with uterine fibroids. F S Rep. Elsevier BV; 2023 Jun;4(2 Suppl):46–50. PMCID: PMC10201292
8. Charkhchi P, Cybulski C, Gronwald J, Wong FO, Narod SA, Akbari MR. CA125 and ovarian cancer: A comprehensive review. Cancers (Basel). MDPI AG; 2020 Dec 11;12(12):3730. PMCID: PMC7763876
9. Pan C, Zhou M, Jian Y, Zeng Y, Wang M, Chen F. CA125: An increasingly promising biomarker of heart failure. Curr Pharm Des. Bentham Science Publishers Ltd.; 2021;27(37):3871–3880. PMID: 33461452
10. Al Mohamady M, Fattah GA, Elkattan E, Bayoumy R, Hamed DA. Correlation of serum CA-125 and progesterone levels with ultrasound markers in the prediction of pregnancy outcome in threatened miscarriage. Int J Fertil Steril. 2016 Jan;9(4):506–511. PMCID: PMC4793171
11. Chen Y, Pan M, Zuo Y, Yang B, Wang S. Research progress of CA125 in endometriosis: Teaching an old dog new tricks. gocm. BMJ; 2022 Dec;2(4):191–198.
12. Akinwunmi BO, Babic A, Vitonis AF, Cramer DW, Titus L, Tworoger SS, Terry KL. Chronic medical conditions and CA125 levels among women without ovarian cancer. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research (AACR); 2018 Dec;27(12):1483–1490. PMCID: PMC6279537
13. Rivo J, Page TF, Arrieta A, Amofah SA, McCann S, Kassaye H, Rodriguez A, Williams ML. The impact of comprehensive pre-visit preparation on patient engagement and quality of care in a population of underserved patients with diabetes: Evidence from the care management medical home center model. Popul Health Manag. Mary Ann Liebert Inc; 2016 Jun;19(3):171–177. PMID: 26440513
14. Vargas E, García-Moreno E, Aghajanova L, Salumets A, Horcajadas JA, Esteban FJ, Altmäe S. The mid-secretory endometrial transcriptomic landscape in endometriosis: a meta-analysis. Hum Reprod Open. Oxford University Press (OUP); 2022 Apr 4;2022(2):hoac016. PMCID: PMC9022214
15. Schwalie PC, Bafligil C, Russeil J, Zachara M, Biocanin M, Alpern D, Aasna E, Deplancke B, Canny G, Goncalves A. Single-cell characterization of menstrual fluid at homeostasis and in endometriosis [Internet]. medRxiv. 2024. p. 2024.05.06.24306766. Available from: https://www.medrxiv.org/content/10.1101/2024.05.06.24306766v1.full-text
16. Dobay A, Haas C, Fucile G, Downey N, Morrison HG, Kratzer A, Arora N. Microbiome-based body fluid identification of samples exposed to indoor conditions. Forensic Sci Int Genet. Elsevier BV; 2019 May;40:105–113. PMID: 30785061
17. Cindrova-Davies T, Zhao X, Elder K, Jones CJP, Moffett A, Burton GJ, Turco MY. Menstrual flow as a non-invasive source of endometrial organoids. Commun Biol. Springer Science and Business

466 Media LLC; 2021 Jun 17;4(1):651. PMCID: PMC8211845
467 18. van der Molen RG, Schutten JHF, van Cranenbroek B, ter Meer M, Donckers J, Scholten RR, van
468 der Heijden OWH, Spaanderman MEA, Joosten I. Menstrual blood closely resembles the uterine
469 immune micro-environment and is clearly distinct from peripheral blood. Hum Reprod. Oxford
470 University Press (OUP); 2014 Feb;29(2):303–314. PMID: 24249743
471 19. Vlasac IM, Stolrow HG, Thayer ZM, Christensen BC, Rivera L. DNA-based cell typing in menstrual
472 effluent identifies cell type variation by sample collection method: toward noninvasive biomarker
473 development for women's health. Epigenetics. Informa UK Limited; 2025 Dec;20(1):2453275. PMID:
474 39825876
475 20. Marečková M, Garcia-Alonso L, Moullet M, Lorenzi V, Petryszak R, Sancho-Serra C, Oszlanczi A,
476 Icoresi Mazzeo C, Wong FCK, Kelava I, Hoffman S, Krassowski M, Garbutt K, Gaitskell K,
477 Yancheva S, Woon EV, Male V, Granne I, Hellner K, Mahbubani KT, Saeb-Parsy K, Lotfollahi M,
478 Prigmore E, Southcombe J, Dragovic RA, Becker CM, Zondervan KT, Vento-Tormo R. An integrated
479 single-cell reference atlas of the human endometrium. Nat Genet [Internet]. 2024 Aug 28; Available
480 from: http://dx.doi.org/10.1038/s41588-024-01873-w PMID: 39198675
481 21. Demir AY, Demol H, Puype M, de Goeij AFPM, Dunselman GAJ, Herrler A, Evers JLH,
482 Vandekerckhove J, Groothuis PG. Proteome analysis of human mesothelial cells during epithelial to
483 mesenchymal transitions induced by shed menstrual effluent. Proteomics. Wiley Online Library;
484 2004 Sep 24;4(9):2608–2623. PMID: 15352236
485 22. Kamiyama S, Teruya Y, Nohara M, Kanazawa K. Impact of detection of bacterial endotoxin in
486 menstrual effluent on the pregnancy rate in in vitro fertilization and embryo transfer. Fertil Steril.
487 Elsevier BV; 2004 Oct;82(4):788–792. PMID: 15482745
488 23. Naseri S, Rosenberg-Hasson Y, Maecker HT, Avrutsky MI, Blumenthal PD. A cross-sectional study
489 comparing the inflammatory profile of menstrual effluent vs. peripheral blood. Health Sci Rep. Wiley
490 Online Library; 2023 Jan 5;6(1):e1038. PMCID: PMC9813904
491 24. Gursoy AY, Caglar GS, Kiseli M, Pabuccu E, Candar T, Demirtas S. CRP at early follicular phase of
492 menstrual cycle can cause misinterpretation for cardiovascular risk assessment. Interv Med Appl
493 Sci. Akademiai Kiado Zrt.; 2015 Dec;7(4):143–146. PMCID: PMC4701164
494 25. Handy AB, Greenfield SF, Yonkers KA, Payne LA. Psychiatric symptoms across the menstrual cycle
495 in adult women: A comprehensive review. Harv Rev Psychiatry. Ovid Technologies (Wolters Kluwer
496 Health); 2022;30(2):100–117. PMCID: PMC8906247
497 26. Xiong B, Yang Y, Fineis FR, Wang JP. DegNorm: normalization of generalized transcript
498 degradation improves accuracy in RNA-seq analysis. Genome Biol. 2019 Apr 16;20(1):75. PMCID:
499 PMC6466807
500 27. Wang L, Nie J, Sicotte H, Li Y, Eckel-Passow JE, Dasari S, Vedell PT, Barman P, Wang L,
501 Weinshiboum R, Jen J, Huang H, Kohli M, Kocher JPA. Measure transcript integrity using RNA-seq
502 data. BMC Bioinformatics. Springer Science and Business Media LLC; 2016 Feb 3;17(1):58.
503 PMCID: PMC4739097
504 28. Byström JW, Vikström L, Rosendal E, Gröning R, Gwon YD, Nilsson E, Sharma A, Espaillat A,
505 Hanke L, McInerney G, Puhar A, Cava F, Karlsson Hedestam GB, Thunberg T, Monsen T, Elgh F,
506 Evander M, Johansson AF, Överby AK, Ahlm C, Normark J, Forsell MN. At-home sampling to meet
507 geographical challenges for serological assessment of SARS-CoV-2 exposure in a rural region of
508 northern Sweden, March to May 2021: a retrospective cohort study. Euro Surveill [Internet].
509 European Centre for Disease Control and Prevention (ECDC); 2023 Mar;28(13). Available from:
510 http://dx.doi.org/10.2807/1560-7917.ES.2023.28.13.2200432 PMCID: PMC10064644
511 29. Zhu Q, Hu Q, Shepherd L, Wang J, Wei L, Morrison CD, Conroy JM, Glenn ST, Davis W, Kwan ML,
512 Ergas IJ, Roh JM, Kushi LH, Ambrosone CB, Liu S, Yao S. The impact of DNA input amount and
513 DNA source on the performance of whole-exome sequencing in cancer epidemiology. Cancer
514 Epidemiol Biomarkers Prev. American Association for Cancer Research (AACR); 2015
515 Aug;24(8):1207–1213. PMCID: PMC4526319
516 30. Schisterman EF, Mumford SL, Sjaarda LA. Failure to consider the menstrual cycle phase may cause
517 misinterpretation of clinical and research findings of cardiometabolic biomarkers in premenopausal
518 women. Epidemiol Rev. Oxford University Press (OUP); 2014;36(1):71–82. PMCID: PMC3873842
519 31. Lamon S, Le Carré J, Petito G, Duong HP, Luthi F, Hiam D, Léger B. The effect of the menstrual
520 cycle on the circulating microRNA pool in human plasma: a pilot study. Hum Reprod. 2023 Jan
521 5;38(1):46–56. PMCID: PMC9825267

522 32. Rocks D, Cham H, Kundakovic M. Why the estrous cycle matters for neuroscience. Biol Sex Differ.
523 2022 Oct 28;13(1):62. PMCID: PMC9615204
524 33. Hmaidan S, Holt E, Finer Z, Session DR, Roach M. The use of endometrial culture for targeted
525 treatment of endometritis in patients experiencing infertility and recurrent pregnancy loss. Fertil
526 Steril. Elsevier BV; 2022 Oct;118(4):e208.
527 34. Muzny CA, Schwebke JR. Asymptomatic bacterial vaginosis: To treat or not to treat? Curr Infect Dis
528 Rep [Internet]. Springer Science and Business Media LLC; 2020 Dec;22(12). Available from:
529 http://dx.doi.org/10.1007/s11908-020-00740-z PMCID: PMC8015387
530 35. Ng BK, Chuah JN, Cheah FC, Mohamed Ismail NA, Tan GC, Wong KK, Lim PS. Maternal and fetal
531 outcomes of pregnant women with bacterial vaginosis. Front Surg. 2023 Feb 13;10:1084867.
532 PMCID: PMC9968788
533 36. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C,
534 Wester K, Hober S, Wernerus H, Björling L, Ponten F. Towards a knowledge-based Human Protein
535 Atlas. Nat Biotechnol. Springer Science and Business Media LLC; 2010 Dec;28(12):1248–1250.
536 PMID: 21139605
537 37. The human proteome in endometrium - The Human Protein Atlas [Internet]. [cited 2025 Apr 16].
538 Available from: https://www.proteinatlas.org/humanproteome/tissue/endometrium

**Table 1: Patient, Cycle, and Tampon level data collection summary**

| A. Patient Level Data for 276 Patients | |
|---|---|
| **Age** | |
| <35 | 166 |
| 35-40 | 74 |
| >40 | 36 |
| **Ethnicity** | |
| Hipanic of Latino | 33 |
| Non-Hispanic | 235 |
| Other/NA | 8 |
| **Race** | |
| Asian | 23 |
| Black or African American | 21 |
| More than on race | 21 |
| White | 207 |
| American Indian or Alaska Native | 1 |
| Other/NA | 3 |
| **Region of Collection** | |
| Midwest | 42 |
| Northeast | 46 |
| South | 91 |
| West | 97 |

| B. Cycle Level Data for 365 Cycles | |
|---|---|
| **Age** | |
| <35 | 218 |
| 35-40 | 99 |
| >40 | 48 |
| **Birth Control** | |
| No birth control | 287 |
| Progestin | 28 |
| Estrogen | 2 |
| Estrogen and Progestin | 41 |
| Other | 7 |
| **Bleeding Phenotype** | |
| Heavy | 170 |
| Medium | 154 |
| Light | 41 |

| C. Tampon/kit level Data for 601 Tampons | |
|---|---|
| **Season of Collection** | |
| Spring | 223 |
| Summer | 157 |
| Autumn | 135 |
| Winter | 86 |
| **Flow Type** | |
| Heavy | 294 |
| Medium | 183 |
| Light | 113 |
| Spotting | 11 |
| **Cycle day** | |
| 1 | 172 |
| 2 | 284 |
| 3 | 145 |

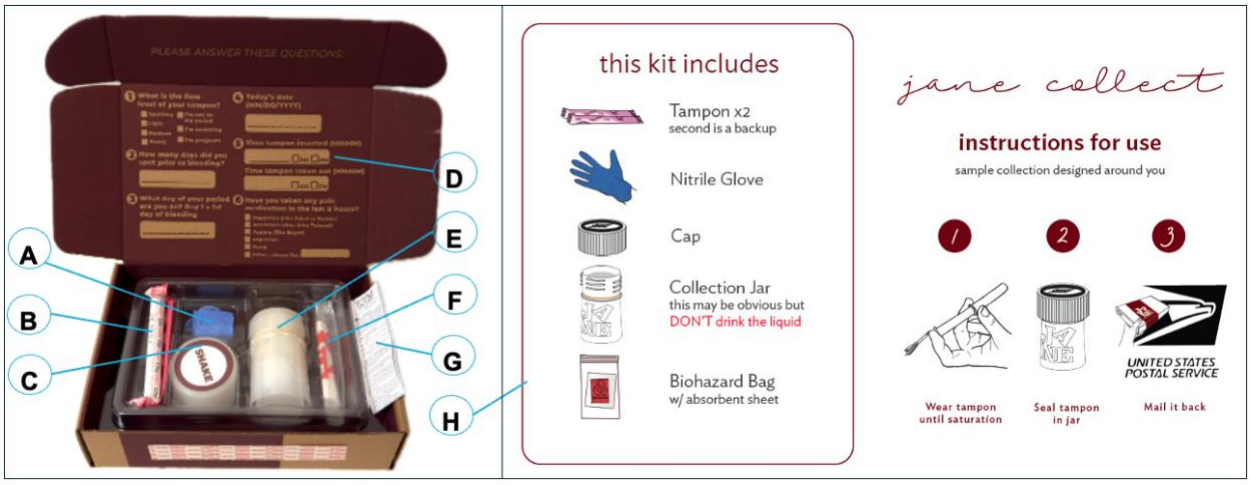
Data was collected at the **a.** patient level, as well as the **b.** cycle and **c.** tampon level. Each patient may have multiple collections within a cycle, and between many monthly cycles. During this time of collection, patients age, change birth control, or may change their bleeding phenotype. We have collected data on these three hierarchical levels to reflect these changes. This dataset includes 601 tampons collected in 365 monthly cycles from 276 participants.

545 **Table 2: Nucleic acid recovery statistics**

| | Nucleic acid recover per 100ul sample lysate | | |
|---|---|---|---|
| | **number of extractions** | **median ng RNA per 100ul lysate** | **average stdev between replicates** |
| | | *ng nucleic acid* | |
| RNA - MagBead | 508 | 251.67 | 28.43 |
| RNA - Column | 627 | 78.20 | 28.14 |
| DNA - MagBead | 12 | 80.80 | NA |

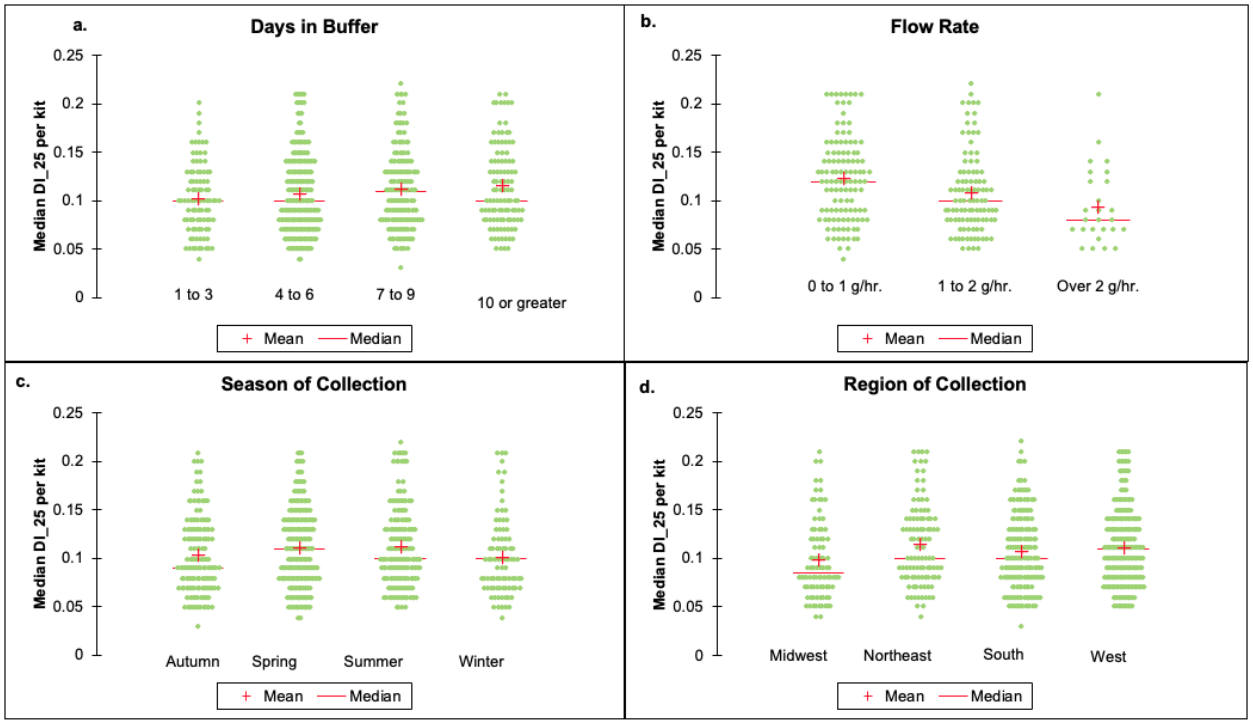
546 Total amount of nucleic acid (in nanograms) per 100ul of lysate was calculated by multiplying the
547 elution volume by the concentration of nucleic acid and then divided by the total volume of sample
548 lysate extracted. Data was calculated only on kits with at least 2 replicates between column-based
549 and magnetic bead-based extraction methods for RNA in order to calculate standard deviation of
550 extracted RNA at the kit level. DNA samples were not extracted in replicates and so standard
551 deviation was not calculated.

552 **Figure 1: NextGen Jane tampon collection kit**




553 NextGen Jane collection kit. **a.** Nitrile glove included for tampon removal. **b.** Included tampon with
554 applicator. **c.** Cap that seals tampon in jar and releases preservative onto tampon. **d.** Basic
555 phenotypic data about current collection is recorded on the inside flap of the box. These include
556 the flow type of the tampon being collected, the day of the cycle, the date, time in and time out for
557 collection, as well as a list of any medications taken during the collection **e.** Collection jar includes
558 a sealed buffer cup with preservative. **f.** Biohazard bag with absorbent material for shipping. **g.**
559 Tampon instructions for use. **h.** Front cover of the tampon collection instructions for use.

**Figure 2: Degradation Index and sequence library size performance metrics**




Sequencing metrics for monitoring degradation of RNA transcripts in menstrual collections. Higher DI scores correlate to more degraded samples. **a.** DI_25 score variance grouped by days sample left in buffer during transit show no significant changes by day (p = 0.107; cohen's f = 0.101). Samples were grouped days 1-3, 4-6, 7-9, and 10 or greater. **b.** DI_25 score plotted by rate of menstrual flow (grams of fluid collected divided by the hours of collections. Samples are bucketed in 0 to 1 grams of fluid lost per hour, 1 to 2 grams lost per hour, and over 2 grams lost per hour. There was a significant decrease in degradation as flow rate increased (p=0.002, cohen's f = 0.230) **c.** & **d.** DI_25 plotted by season (U.S) and by region (U.S.) the collection occurred. Midwest collected samples showed a lower mean DI_25 compared to other regions (p = 0.023; cohen's f = 0.127). Analysis was performed at the individual sequence level (n=1,135)

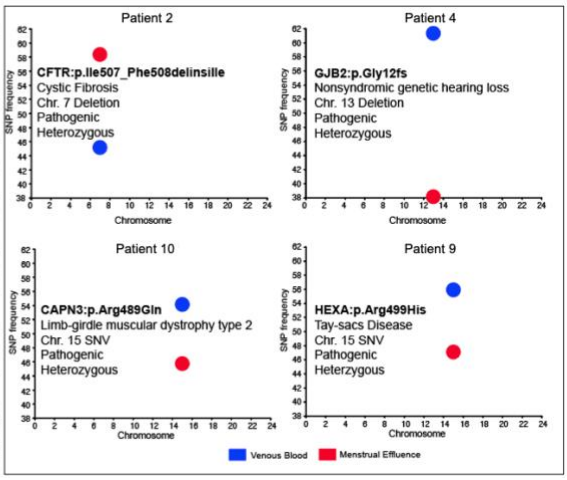
571 **Figure 3: SNV detection in DNA from matched whole blood and menstrual effluence**

572 **samples**

**a.**

| Patient ID | SNVs detected in Venous | SNVs detected in Menstrual | Rate of Venous SNV confirmed in Menstrual |
|---|---|---|---|
| Patient 1 | 2642 | 2684 | 0.9637 |
| Patient 2 | 2663 | 2787 | 0.9884 |
| Patient 3 | 2713 | 3080 | 0.9849 |
| Patient 4 | 2616 | 2704 | 0.9824 |
| Patient 5 | 2656 | 2890 | 0.9857 |
| Patient 6 | 2590 | 2608 | 0.9842 |
| Patient 7 | 2641 | 2718 | 0.9909 |
| Patient 8 | 2701 | 2926 | 0.9907 |
| Patient 9 | 2804 | 3057 | 0.9904 |
| Patient 10 | 3194 | 3214 | 0.9759 |
| Patient 11 | 2603 | 2771 | 0.9697 |
| Patient 12 | 2542 | 2648 | 0.9898 |
| **Average** | **2697** | **2841** | **0.9830** |
| **Standard Deviation** | **170** | **193** | **0.0089** |

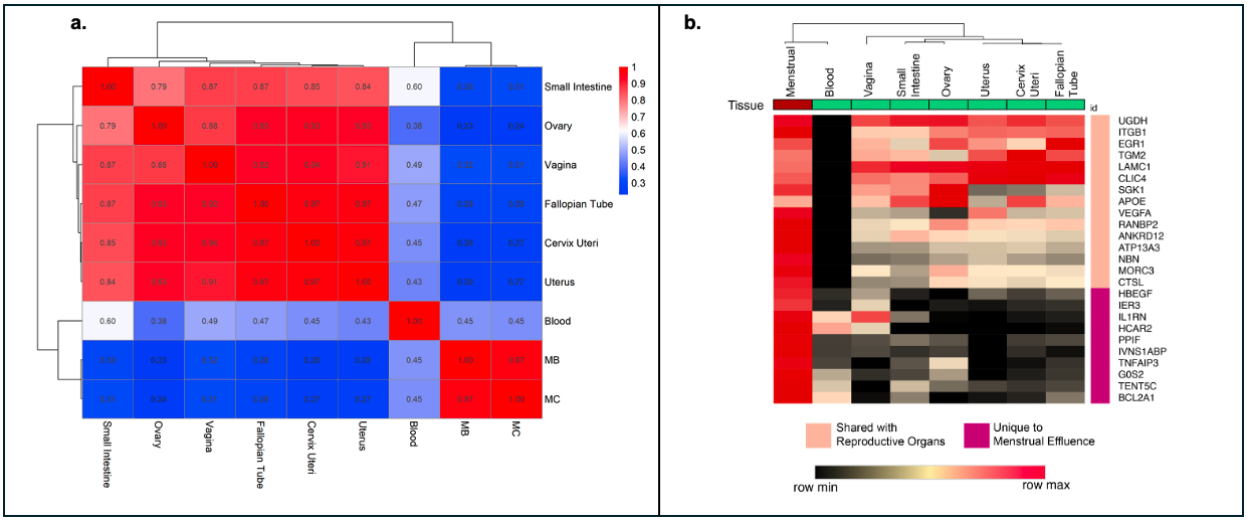
**b.**




573

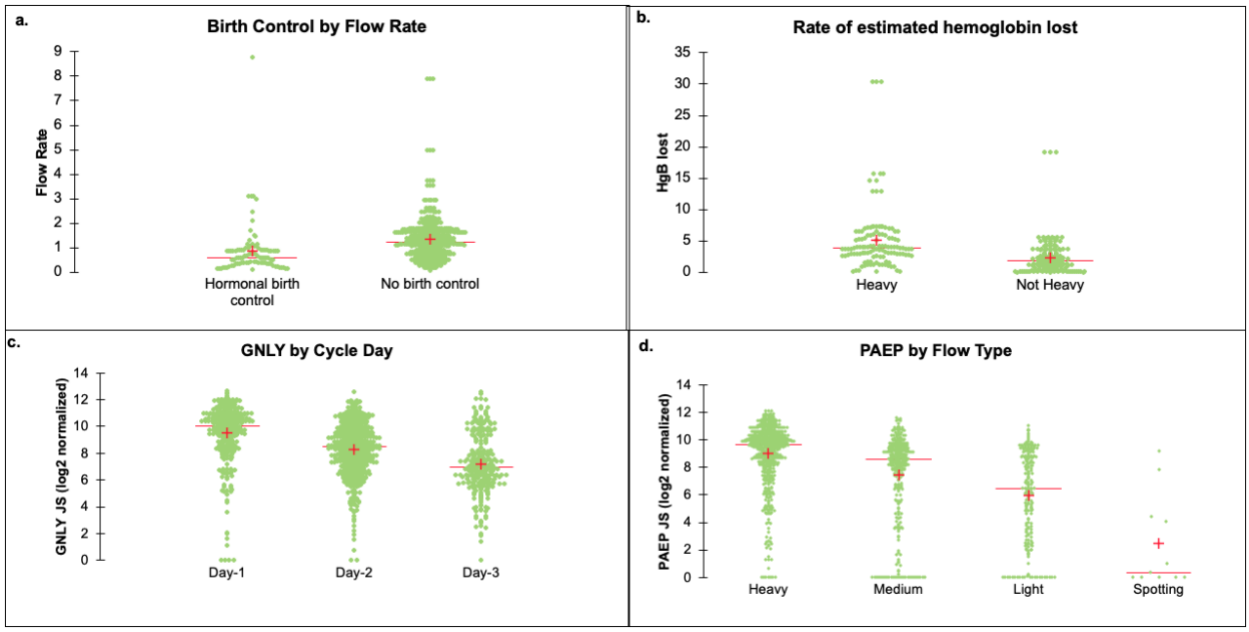
574 SNV summary statistics for the comparison of SNVs between 12 patients who were exome
575 sequenced using DNA extracted from menstrual effluence and from a venous blood draw. **a.**
576 Table summarizing the number of SNVs recorded in the venous blood and menstrual effluence,
577 and the rate of venous blood SNVs that were confirmed also in menstrual effluence. **b.** Four
578 patients with known pathogenic mutations in CFTR, GJB2, CAPN3 and HEXA were confirmed
579 present in both menstrual effluence and venous blood. Charts show mutation plotted against the
580 frequency of the SNV detected. Red points are menstrual and blue points are venous. 24 samples
581 from 12 participants were used in this analysis.

**Figure 4: Sample type comparisons in menstrual effluence compared to other tissue types**




Menstrual effluence gene expression global differences to other tissue types. **a.** Spearman correlation matrix of global RNA transcriptional activity in menstrual effluence (MB: tampon; MC: menstrual cup) compared to public datasets for small intestine, ovary, vagina, fallopian tube, cervix, uterus and blood. Red indicates a high level of transcriptional correlation between tissue types, while white and blue indicate a low level of correlation to other tissue types. 4,235 samples from GTeX and 218 samples from the NGJ dataset were used for this analysis. **b.** Top 25 differentially expressed genes between menstrual effluence and venous blood. Median gene values for 20 sequences from each GTeX tissue type were taken for comparison. Shared genes are grouped between reproductive tissue and menstrual (peach) vs menstrual samples with little to now expression in GTeX tissues (magenta). Expression differences in KRT13 between menstrual effluence collected in a menstrual cup or a tampon ($p < 0.0001$; cohen's $f = 2.072$). Analysis was done at the sequencing level (ESG) for 70 menstrual cup samples and 68 tampon samples.

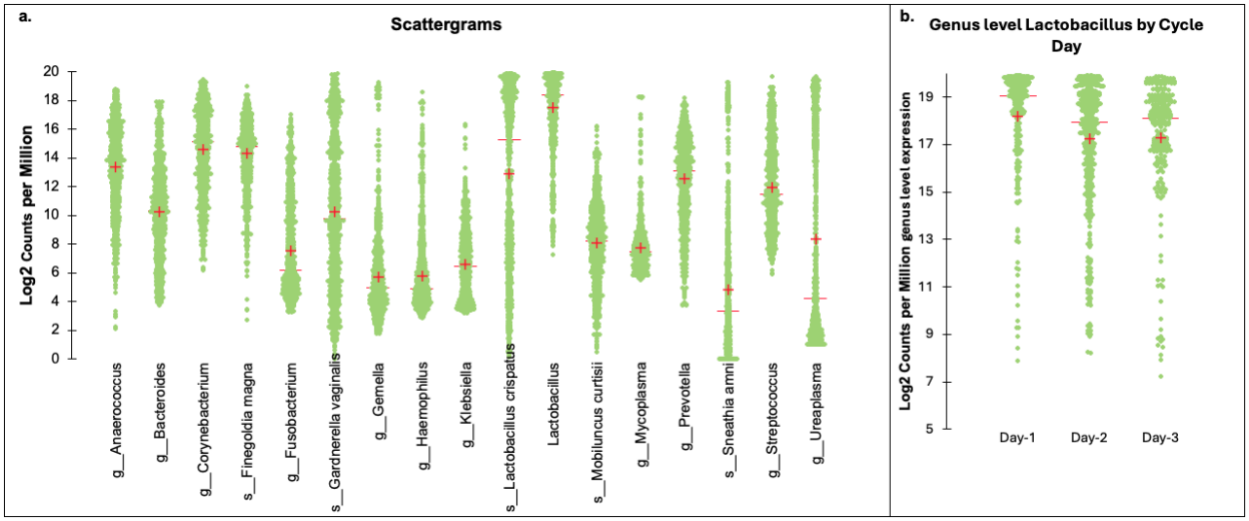
597 **Figure 5: Flow rate and gene expression differences by sample collection characteristics**




598

599 Flow rate and hemoglobin content by birth control and heavy menstrual bleeding and the
600 differential expression of endometrial genes by cycle day. **a.** Flow rate calculated as grams of
601 fluid collected per hour plotted against patient use of hormonal birth control. Patients on any form
602 of hormonal birth control show lower flow rates ($p < 0.0001$; cohen's $f = 0.204$). Analysis was
603 performed at the kit level for 241 tampons with recorded flow rate. **b.** Flow rate multiplied by the
604 rfu (relative fluorescence units) spectrophotometer reading as a metric for hemoglobin within the
605 sample charted by whether patients self-report as heavy menstrual bleeders or normal menstrual
606 bleeders. Heavy bleeders show an increase in lost rate of hemoglobin per hour ($p < 0.0001$;
607 cohen's $f = 0.352$). Analysis was performed at the kit level for 241 tampons with recorded flow
608 rate. **c.** GNLY expression, an endometrium-enriched gene, by the day in the cycle (first cycle day
609 represents first day of bleeding), shows significant expression changes by cycle day ($p < 0.0001$;
610 cohen's $f = 0.394$). **d.** PAEP, another endometrium enriched gene charted by the day specific flow
611 type during the tampon collection ($p < 0.0001$; cohen's $f = 0.469$). Mean and Median values are
612 represented by a red "+" and "—", respectively. C and d analyses were performed at the sequence
613 level (ESG) for 1,135 samples.

614 **Figure 6: RNA transcriptional activity of key bacteria in menstrual effluence**




615

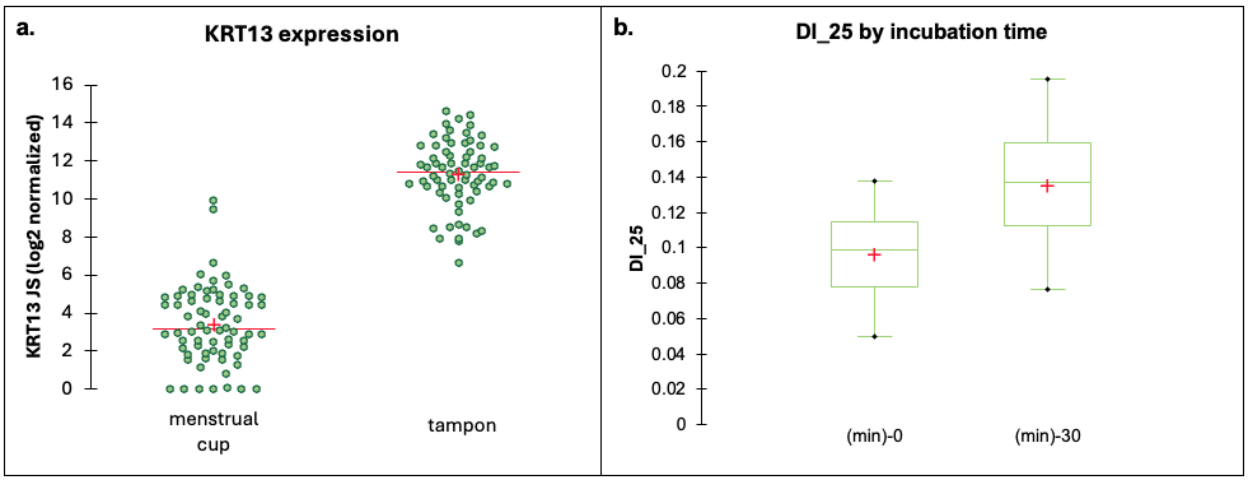
616 **a.** *L. crispatus*, which is one beneficial lactobacillus species in the vaginal cavity alongside BV
617 causing bacteria at the genus level. CPM counts per genus were summed before log2
618 transformation. Mean and Median values are represented by a "+" and "—", respectively. **b.**
619 Genus level *Lactobacillus* shows a significant decrease in transcriptional activity on day 2 and
620 day 3 of menstruation, compared to day 1 ($p < 0.0001$, cohen's $f = 0.171$). Analyses were
621 performed at the sequence level (ESG) for 1,135 samples.

622 **Supplementary Table 1: Sample filtering parameters**

623 Menstrual samples collected on cycle days 1-3 that had complete metadata for collection date,
624 bleeding phenotype, flow type, birth control, collection time and reproductive stage were included
625 in the analysis. Samples were further filtered by technical sequencing QC metrics including proper
626 strandedness of reads, ribosomal rRNA contamination and uniquely mapped reads.

| | |
|---|---|
| **All Samples** | **1867** |
| **Cycle day and Menstrual** | **1600** |
| Cycle days 1-3 | |
| Exclude vaginal only samples | |
| **Missing Data (0, Not Recorded, or NA)** | **1346** |
| Collection Date | |
| Bleeding Phenotype | |
| Flow type | |
| Birth Control | |
| Collection time | |
| Reproductive Stage | |
| **Technical Filters** | **1135** |
| Strandedness >0.8 | |
| rRNA < 3% | |
| Uniquely mapped reads >8M | |

635 **Supplemental Figure 1: Supplemental analyses on menstrual cups**




636

637 **a.** Expression differences in KRT13 between menstrual effluence collected in a menstrual cup or
638 a tampon ($p < 0.0001$; cohen's $f = 2.072$). Analysis was done at the sequencing level (ESG) for 70
639 menstrual cup samples and 68 tampon samples. **b.** Menstrual effluence collected in menstrual
640 cups was either directly put into preservation buffer after collection or incubated at 37oC for 30
641 minutes before preservation ($p=0.003$; cohen's $f = 0.606$). A total of 10 diva cups from 5
642 participants were used in this analysis.