The vagino-cervical microbiome as a woman's life history

Zhuye Jie[1,2,3,*], Chen Chen[1,3,4,*,†], Lilan Hao[1,3], Fei Li[1,3], Liu Tian[1,3], Xin Tong[1,3], Liju Song[1,3], Kaiye Cai[1], Xinlei Yu[1,3], Ying Li[1,3], Weiqiao Rao[1], Yunli Liao[1], Dongsheng Zhou[1], Xuemei Qiu[1], Xiaomin Chen[1], Jin Zi[1], Na Liu[1], Wei Zhang[1], Xiao Liu[1], Liang Xiao[1,5,6], Xun Xu[1,3], Huanming Yang[1,7], Jian Wang[2,7], Yong Hou[1], Karsten Kristiansen[1,4,6], Huijue Jia[1,2,3,7†]

1. BGI-Shenzhen, Shenzhen 518083, China;
2. Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, BGI-Shenzhen, Shenzhen, China;
3. China National Genebank, BGI-Shenzhen, Shenzhen 518120, China;
4. Department of Biology, Ole MaalØes Vej 5, University of Copenhagen, Copenhagen, Denmark;
5. Shenzhen Engineering Laboratory of Detection and Intervention of human intestinal microbiome, BGI-Shenzhen, Shenzhen, China;
6. BGI-Qingdao, BGI-Shenzhen, Qingdao, 266555, China;
7. James D. Watson Institute of Genome Sciences, Hangzhou, China
8. Macau University of Science and Technology, Avenida Wai long, Taipa, Macau, China.

* These authors contributed equally to this work.
† To whom correspondence should be addressed: H.J., jiahuijue@genomics.cn or C.C., chenchen1@genomics.cn

The gut microbiome has been the center of attention for human commensal microbiome studies. The vaginal microbiome is also densely populated with bacteria, viruses and fungi, and the presence of microorganisms beyond the cervix is increasingly reported in non-infectious conditions[1–3]. Due to the over 90% of human sequences in female reproductive tract samples[3,4], metagenomic information has been very limited. 16S rRNA gene amplicon sequencing studies have identified community types in the vaginal microbiota, and observed its dynamic changes due to menstrual cycles and sexual behaviors in small cohorts[5,6]. Here we perform metagenomic shotgun sequencing on cervical samples from 516 women of reproductive age (more than 10-fold of the Human Microbiome Project (HMP)[4]), and dissect major factors, especially pregnancy and delivery histories and contraception methods on the microbiome composition. Features of other body sites, such as mood fluctuations and facial speckles could potentially be deduced from the vagino-cervical microbiome. Our results offer an unprecedented glimpse into the microbiota in the female reproductive tract and imply disease susceptibilities that may be relieved by behavioural changes.

Dominant bacterial and non-bacterial members of the vagino-cervical metagenome

To explore the vagino-cervical microbiome, 516 healthy Chinese women aged 21-52 (median 29, 95% CI: 23–39) were recruited during a physical examination (**Sup. Data 1**). Cervical samples were collected using a cytobrush and immediately preserved on Flinders Technology Associates (FTA) cards. Metagenomic shotgun sequencing was

performed on the DNA samples, and high-quality non-human reads were used for taxonomic profiling of the vagino-cervical microbiome (**Fig. 1**).

In agreement with 16S rRNA gene amplicon sequencing data from the US[5], the vagino-cervical microbiota of this Asian cohort was mostly Lactobacilli-dominated, while 19.19% of women harboured over 50% *Gardnerella vaginalis* (**Fig. 1a, Extended Data Fig. 1**). Some of the volunteers were dominated by *Atopobium vaginae* but not *G. vaginalis*, which were commonly believed to co-occur in bacterial vaginosis (BV)[7]. 22.67% of the cohort was dominated by *Lactobacillus iners* (**Fig. 1a, Extended Data Fig. 1**), which was far less protective against bacterial and viral infections compared to *Lactobacillus crispatus*. Rare subtypes such as *Bifidobacterium breve, Lactobacillus johnsonii*, *Streptococcus anginosis*, *Lactobacillus amylovorus* and *Escherichia coli* were also detected in this cohort (**Fig. 1a**). Other microorganisms including *Clamydia trachomatis*, *Ureaplasma parvum*, human papillomavirus (HPV), herpesviruses, *Haemophilus influenzae* and Enterobacteria phage were abundant in some individuals (**Fig. 1**). According to the metagenomic data, the vagino-cervical microbiome varied widely among individuals (**Fig. 1a, Extended Data Fig. 1a**), and the mean proportion of non-bacterial sequences was 3.45%. The vaginal types of *G. vaginalis*, *L. crispatus*, and *B. breve* showed lower proportions of human sequences than other types (**Extended Data Fig. 1b**).

Factors shaping the vagino-cervical microbiome and beyond

We computed the prediction value (5-fold cross-validated random forest model) of female life history questionnaire on the microbiome data, and found the most important factors to be pregnancy, marriage, mode of delivery, age at vaginal sexual debut, actively breast-feeding, contraceptive methods, menstrual cycle, and sexual intercourse frequency (**Fig. 2a**).

Weaker but nonetheless significant dependences were observed between the vagino-cervical microbiome and practices such as sexual intercourse within 24 hours, HPV vaccination, vaginal douching, oral probiotics, antibiotics and herbal medicine (**Fig. 2a, Sup. Data 2**).

As a whole, these questionnaire data collected for the female reproductive tract samples showed the greatest explained variances for the vagino-cervical microbial composition, followed by other data collected on the same day, such as faecal microbiome composition, psychological questionnaire, plasma metabolites, immune indices, facial skin imaging and medical test data (**Fig 2b, Extended Data Fig. 2**). For instance, self-assessed depression was the strongest factor in the psychological questionnaire to predict vagino-cervical microbiome composition, spots on the cheeks constituted the strongest factor in facial skin measurements, and plasma phosphoserine was the most important among metabolites (**Extended Data Fig. 2**).

Specific influences from pregnancy histories and contraception

Marriage was one of the most significant factors to influence the vagino-cervical microbiome (**Fig. 2a**). It showed negative correlations with relative abundances of *L. crispatus*, *Acinetobacter* species, *L. jensenii*, *L. vaginalis*, *Ureaplasma parvum* and *Comamonas testosteroni*, and positive correlation with *Bifidobacterium breve* (**Fig 3a, Sup. Data 2**). Women who got married had higher concentration of vitamin D, and their plasma testosterone (responsible for sexual drive), dehydroepiandrosterone and creatinine had declined (**Fig 3a**). The age at vaginal sexual debut showed positive correlations with *Gardnerella vaginalis* and *Scardovia*, and negative associations with *L. crispatus* and *L. iners* (**Fig 3a, Sup. Data 2)**.

Similarly, the women who went through pregnancy showed decreased *Lactobacillus*, including *L. crispatus*, *L. vaginalis*, and *L. jensenii*, but were enriched for Actinobacteria, Bifidobacteriaceae, Anelloviridae and Torque teno virus (the most abundant anellovirus in the human virome) in the vagino-cervical microbiome (**Fig. 3c, Sup. Data 2, Extended Data Fig.3**). We observed increased concentrations of vitamin D, blood glucose and direct bilirubin in women with previous pregnancy, but lower concentrations of testosterone, androstenedione, dehydroepiandrosterone and methionine (**Fig. 3c**). Vaginal deliveries were associated with decreased *L. crispatus*, *L. jensenii*, *Bartonella* species, *Prevotella timonensis*, and increased *Lactobacillus* sp. 7_1_47FAA (**Fig. 3d, Sup. Data 2**). We observed a lower relative abundance of *Ureaplasma parvum* in individuals who delivered by cesarean section (**Fig. 3e, Sup. Data 2**), a bacterium commonly isolated from pregnant women[8] and recently reported in the lower respiratory tract of preterm infants[9]. Compared with vaginal deliveries, the concentrations of

testosterone, androstenedione, methionine, threonine and tryptophan were significantly lower in women who experienced caesarean section (**Fig 4d,3e**). In addition, these women more often suffered from an abnormal leucorrhoea (**Fig 3e**).

Among our cohort, 63 volunteers happened to be actively breast-feeding. *L. crispatus*, and *L. iners* was found to be less abundant in these actively breast-feeding individuals (**Fig. 3f)**. Alkaline phosphatase, vitamin D and lead (Pb) were found to be more abundant in the breast-feeding individuals, while progesterone and creatine levels were slightly reduced (**Fig 3f**). Gravida and para, especially deliveries by caesarean section, appeared associated with facial skin problems such as UV spots and porphyrins, but these problems may be reduced by breast-feeding (**Fig 3c-3f**).

Integrated association network using a wisdom of crowds approach[10] also revealed interesting patterns (Spearman's correlation, random forest and linear regression, **Extended Data Fig. 4**). The relative abundance of *G. vaginalis* was positively associated with a poor mental state, poor sleep and appetite (**Extended Data Fig. 4**), possibly due to reduced GABA (γ-amino butyric acid) production from Lactobacilli[11], and the results advocate for getting a partner at a younger age (**Fig. 3**). *L. iners* negatively correlated with times of spontaneous abortion, days after last menses, condom usage and the concentration of vitamin D and vitamin A, and positively correlated with plasma concentrations of asparagine and alanine. In addition, vaccination history, such as these against HPV, tetanus toxoid, and bacillus Calmette-Guerin, may be associated with higher relative abundances of *L. iners*.

We found multiple significant associations between the contraceptive methods of participants and their vagino-cervical microbiome. Condom usage showed negative correlations with *L. iners* and *Comamonas* species, and positive associations with *Prevotella* species (**Sup. Data 2, Extended Data Fig. 4**). Oral contraceptives, on the other hand, were associated with increased *Ureaplasma parvum*, *Comamonas* species and Mycoplasmataceae (**Fig. 4g, Sup. Data 2, and Extended Data Fig. 4**). *Comamonas* has been implicated in fecundity in *C. elegans* and identified as a marker for infertility due to endometriosis [2,12]. Moreover, oral contraceptives exhibited a positive correlation with plasma homocitrulline, and a negative correlation with good physical shape (such as waist-to-hip ratio and body weight), consistent with decline in butyrate-producing faecal microbiota species such as *Faecalibacterium prausnizi* and *Roseburia intestinalis* (**Fig. 3g**). Thus, none of the contraceptive methods appeared healthy for the microbiome.

Menstrual phases are known to influence the vaginal microbiota[2,6]. We confirmed in our cohort that *L. iners* gradually decreased in proportion after menses, coinciding with the increase in progesterone and the decrease of plasma threonine and arginine (**Fig 3h**). White blood cell counts (WBC) also recovered after menses, consistent with greatest susceptibility to infection at the onset of menses. *L. vaginalis*, *L. johnsonii* and *Weissella* species showed negative correlations with self-reported regular periods (**Extended Data Fig. 4**). *Streptococcus agalactiae*, *Propionibacterium acnes* and *Dialister micraerophilus* were positively associated with a heavier flow (**Extended Data Fig. 4**).

Many women experience dysmenorrhea during menses. Dysmenorrhea enriched for Pseudomonadales, *Acinetobacter* and Moraxellaceae, while lower in plasma level of histidine (**Fig 3i**), consistent with these bacteria producing histidine decarboxylases to convert histidine into histamine[13].

We also found some significant omics factors varied among the different vaginal types (**Extended Data Fig. 5**). The women with types of *B. breve* and *G. vaginalis* had older age at vaginal sexual debut (**Extended Data Fig. 5a**). The types of *L. crispatus* and *L. jensenii* were overrepresented in the women who were not married yet or had fewer pregnancies (**Extended Data Fig. 5b, c**). We observed more abundant of proline and creatinine in the *S. anginosus*-type individuals, but lower concentrations of 11-deoxycortisol and 17α-hydroxyprogesterone (**Extended Data Fig. 5d-g**). The concentrations of 11-deoxycortisol, 17α-hydroxyprogesterone and testosterone were higher in *L. crispatus*-type than *L. iners* and *G. vaginalis* types (**Extended Data Fig. 5f-h**). Vaginal types also appeared associated with facial skin features such as *A. vaginae* and *L. johnsonii* types had higher rates of skin problem than other types (**Extended Data Fig. 5i-l**). Faecal *Neisseria* species, *Haemophilus* species and *Aggregatibacter* species showed greater abundance in *L. jensenii*-type (**Extended Data Fig. 5m-o**).

While lacking significant association with the vagino-cervical microbial composition, the immune repertoire data showed multiple associations with functional pathways in the vagino-cervical microbiome, such as purine and pyrimidine metabolism, synthesis of branched chain amino acids, histidine and arginine (**Extended Data Fig. 6**). Red blood

cell counts, vitamin A and hydroxyl vitamin D levels negatively associated with CDP-diacylglycerol biosynthesis pathways, consistent with presence of diacylglycerol kinase in Lactobacilli with anti-inflammatory functions[14,15]. Fecal *Coprococcus comes*, a bacterium previously reported to associate with cytokine response to *Candida albicans*[16], was seen here to associate with isoleucine pathways in the vagino-cervical microbiome (**Extended Data Fig. 6**).

Overall, our comprehensive data on a healthy cohort of young women could help target efforts aimed at promoting a healthy reproductive tract microbiota and minimizing preterm birth, as well as preventing infections from viruses such as the human immunodeficiency virus (HIV)[17] and HPV. For example, whether vitamin D supplementation should be considered in African countries to reduce vaginal *L. iners*, and therefore HIV and BV susceptibility. While the focus in the field has always been infection, our data highlight major aspects (e.g. marriage, delivery methods, breast-feeding, contraception) that are worth further investigations for women in the modern world. As a densely populated microbiota other than the distal gut, the vagino-cervical microbiota also has the potential to reflect or even influence physiology elsewhere in the human body.

Methods Summary

Informed consent was obtained from all 516 healthy women from the Institutional Review Boards (IRB) at BGI-Shenzhen before participation in the study. Cervical samples and faecal samples were collected from all volunteers during physical

examination. DNA was extracted as described[2,18], singled-end libraries were constructed and sequenced using the BGISEQ-500 platform[19]. Medical test data were performed by the physical examination center. Blood was drawn from a cubital vein of volunteers, then carried out a series of experiments, including plasma metabolites (hormone, vitamin, amino acid and microelement) and immune indices. Facial skin characteristics were evaluation by VISIA-CR™ imaging system. Physical fitness tests were evaluation by a series of HK6800 testers. Three kinds of online questionnaires about female life history questionnaire, lifestyle questionnaire and psychological questionnaire were filled in by volunteers.

Accession codes

Metagenomic sequencing data for all samples have been deposited to the (CNGB) database under the accession code CNP0000287.

Acknowledgments

This research was supported by the Shenzhen Municipal Government of China (JCYJ20160229172757249 and JCYJ20170817145523036) and Shenzhen Peacock Plan (No. KQTD20150330171505310). The authors are very grateful to colleagues at BGI-Shenzhen for sample collection, DNA extraction, library construction, sequencing, and discussions. We thank Dr. Xin Mu and Dr. Kou Qin for helpful comments.

Author contributions

H.J. and C.C. conceived and organized this study. J.W. initiated the overall health project. C.C., L.H., F.L, L.S., Y.L., X.Y. and N.L. performed the sample collection and

questionnaire collection. X.T., Y.K., W.R., Y.L., D.Z., X.Q., X.C., and J.Z. provided the omics data. Z.J., C.C., F.L, L.H., and L.T. performed the bioinformatic analyses, H.J., C.C., and Z.J. wrote the manuscript. All authors contributed to data and texts in this manuscript.

Competing financial interests

The authors declare no competing financial interest.

References

1. Aagaard, K. *et al.* The placenta harbors a unique microbiome. *Sci. Transl. Med.* **6,** 237ra65 (2014).
2. Chen, C. *et al.* The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat. Commun.* **8,** 875 (2017).
3. Li, F. *et al.* The metagenome of the female upper reproductive tract. *Gigascience* **7,** (2018).
4. Methé, B. A. *et al.* A framework for human microbiome research. *Nature* **486,** 215–21 (2012).
5. Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci.* **108,** 4680–7 (2010).
6. Gajer, P. *et al.* Temporal Dynamics of the Human Vaginal Microbiota. *Sci. Transl. Med.* **4,** 132ra52-132ra52 (2012).
7. Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *N. Engl. J. Med.* **353,** 1899–1911 (2005).
8. Anderson, B. L. *et al.* Pregnancy-induced changes in immune protection of the genital tract: defining normal. *Am. J. Obstet. Gynecol.* **208,** 321.e1-9 (2013).
9. Pattaroni, C. *et al.* Early-Life Formation of the Microbial and Immunological Environment of the Human Airways. *Cell Host Microbe* **24,** 857–865.e4 (2018).
10. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9,** 796–804 (2012).
11. Strandwitz, P. *et al.* GABA-modulating bacteria of the human gut microbiota. *Nat. Microbiol.* (2018). doi:10.1038/s41564-018-0307-3

12. MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J. & Walhout, A. J. M. Diet-induced developmental acceleration independent of TOR and insulin in C. elegans. *Cell* **153,** 240–52 (2013).

13. Cundell, D. R., Devalia, J. L., Wilks, M., Tabaqchali, S. & Davies, R. J. Histidine decarboxylases from bacteria that colonise the human respiratory tract. *J. Med. Microbiol.* **35,** 363–6 (1991).

14. Ganesh, B. P. *et al.* Diacylglycerol kinase synthesized by commensal Lactobacillus reuteri diminishes protein kinase C phosphorylation and histamine-mediated signaling in the mammalian intestinal epithelium. *Mucosal Immunol.* (2017). doi:10.1038/mi.2017.58

15. Andrada, E., Liébana, R. & Merida, I. Diacylglycerol Kinase ζ Limits Cytokine-dependent Expansion of CD8+ T Cells with Broad Antitumor Capacity. *EBioMedicine* **19,** 39–48 (2017).

16. Schirmer, M. *et al.* Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* **167,** 1125–1136.e8 (2016).

17. Klatt, N. R. *et al.* Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science (80-.).* **356,** 938–945 (2017).

18. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).

19. Fang, C. *et al.* Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *Gigascience* **7,** (2018).

20. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. *Genetics* **201,** 459–472 (2015).

21. Segata, N. MetaPhlan2. *Nat Methods* (2012).

22. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32,** 834–841 (2014).

23. Wang, J. & Jia, H. Metagenome-wide association studies: fine-mining the microbiome. *Nat. Rev. Microbiol.* **14,** 508–522 (2016).

24. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32,** 822–828 (2014).

25. Moll1, J. M. *et al.* Pathway-based approach to classify unannotated gut bacteria associated with insulin sensitivity. *Under Revis. Nat Microbiol* (2019).

26. Louppe, G., Wehenkel, L., Sutera, A. & Geurts, P. in *Advances in Neural Information Processing Systems 26* (eds. Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. Q.) 431–439 (Curran Associates, Inc., 2013). at <http://papers.nips.cc/paper/4928-understanding-variable-importances-in-forests-of-randomized-trees.pdf>

Figure Captions

Figure 1 Vagino-cervical microbiome of the cohort. The microbial composition in each sample at the species level according to MetaPhylAn2 is shown. The samples are ordered according to hierachical clustering. Taxa names in black and blue denote bacteria and viruses, respectively.

Figure 2 Global view of factors influencing the vagino-cervical microbiome. a, Female life history questionnaire entries on the vagino-cervical microbiome, ordered according to their 5-fold cross-validated random forest (RFCV) importance on the microbiome composition. **b**, Predicting the vagino-cervical microbiome from each omics and vice versa using stepwise redundancy analysis. Numbers on the straight arrows indicate adjusted R-squared from omics data to vagino-cervical microbiome; numbers on the curved arrows indicate adjusted R-squared from vagino-cervical microbiome to omics data.

Figure 3 Specific influences from reproductive factors on omics data. Effect size of the factors shown (**a**, Marriage; **b**, Age at vaginal sexual debut; **c**, Pregnancy history; **d**, Number of vaginal deliveries; **e**, Number of cesarean deliveries; **f**, Current breast-feeding; **g**, Oral contraceptives; **h**, Days since last menstrual bleeding; **i**, Dysmenorrhea). Wisdom of crowds was performed to calculate the correlation of omics data. The bars are colored according to the type of omics, as in **Fig. 2b**. The length of the bars represents Spearman's correlation coefficient between the actual measurements and the values predicted by 5-fold cross-validated random forest models.

A plus sign denotes Q-value <0.1, an asterisk denotes Q-value <0.05, two asterisks denote Q-value <0.01, three asterisks denote Q-value <0.001, four asterisks denote Q-value <0.0001, and five asterisks denote Q-value <0.00001.

Online Methods

Study Cohort

All the 516 volunteers were recruited between May 2017 and July 2017 during physical examination in BGI-Shenzhen, China. Baseline characteristics of the cohort are shown in **Sup. Data 1**.

The study was approved by the Institutional Review Boards (IRB) at BGI-Shenzhen, and all participants provided written informed consent at enrolment.

Demographic Data Collection

During physical examination, the volunteers received three kinds of online questionnaire. The female life history questionnaire contained pregnancy and delivery history, menstrual phases, sexual activity, contraceptive methods and so on. The lifestyle questionnaire contained age, disease history, eating and exercise habits and so on. The psychological questionnaire contained the evaluation of irritability, dizziness, frustration, fear, appetite, self-confidence, resilience and so on (**Sup. Data 1**).

Samples Collection

Cervical samples were collected and smeared in the Flinders Technology Associates (FTA) cards by the doctor during gynecological examination. Fecal samples were self-collected by volunteers. Cervical samples and faecal samples were stored for metagenomic shotgun sequencing. The overnight fasting blood samples were drawn from a cubital vein of volunteers by the doctors.

DNA extraction and metagenomics shotgun sequencing

DNA extraction of cervical samples and fecal samples was performed as described[2,18], respectively. Metagenomic sequencing was done on the BGISEQ-500 platform (50bp of singled-end reads for cervical samples and four libraries were constructed for each lane; 100bp of singled-end reads for fecal samples and four libraries were constructed for each lane)[19].

Amino Acid Measurements

40 µl plasma was deproteinized with 20 µl 10% (w/v) sulfosalicylic acid (Sigma) containing internal standards, then 120 µl aqueous solution was added. After centrifuged, the supernatant was used for analysis. The analysis was performed by ultra high pressure liquid chromatography (UHPLC) coupled to an AB Sciex Qtrap 5500 mass spectrometry (AB Sciex, US) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.8 µm, 2.1 × 100 mm) was used for amino compound separation with a flow rate at 0.5 ml/min and column temperature of 55 °C. The mobile phases were (A) water containing 0.05% and 0.1% formic acid (v/v), (B) acetonitrile containing 0.05% and 0.1% formic acid (v/v). The gradient elution was 2% B kept for 0.5 min, then changed linearly to 10% B during 1 min, continued up to 35% B in 2 min, increased to 95% B in 0.1 min and maintained for 1.4 min. Multiple Reaction Monitoring (MRM) was used to monitor all amino compounds. The mass parameters were as follows, Curtain gas flow 35 L/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 60 l/min, Ion Source Gas 2 (GS

2) flow 60 l/min, IonSpray Voltage (IS) 5500V, temperature 600 °C. All amino compound standards were purchased from sigma and Toronto research chemical (TRC).

Hormone Measurements

250 µl plasma was diluted with 205 µl aqueous solution, For SPE experiments, HLB (Waters, USA) was activated with 1.0 ml of dichloromethane, acetonitrile, methanol, respectively and was equilibrated with 1.0 ml of water. The pretreated plasma sample was loaded onto the cartridge and was extracted using gravity. Clean up was accomplished by washing the cartridges with 1.0 ml of 25% methanol in water. After drying under vacuum, samples on the cartridges were eluted with 1.0 ml of dichloromethane. The eluted extract was dried under nitrogen and the residual was dissolved with 25% methanol in water and was transferred to an autosampler vial prior to LC–MS/MS analysis. The analysis was performed by UHPLC coupled to an AB Sciex Qtrap 5500 mass spectrometry (AB Sciex, US) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Phenomone Kinetex C18 column (2.6 µm, 2.1 × 50 mm) was used for steroid hormone separation with a flow rate at 0.8 ml/min and column temperature of 55 °C. The mobile phases were (A) water containing 1mM Ammonium acetate, (B) Methanol containing 1mM Ammonium acetate. The gradient elution was 25% B kept for 0.9min, then changed linearly to 40% B during 0.9min, continued up to 70% B in 2 min, increased to 95% B in 0.1 min and maintained for 1.6 min. Multiple Reaction Monitoring (MRM) was used to monitor all steroid hormone compounds. The mass parameters were as follows, Curtain gas flow 35 l/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 60 l/min, Ion Source

Gas 2 (GS 2) flow 60 l/min, Nebulizer Current (NC) 5, temperature 500 °C. All steroid hormone profiling compound standards were purchased from sigma, Toronto research chemical (TRC), Cerilliant and DR. Ehrenstorfer.

Microelement Measurements

200 µl of whole blood were accurately transferred into a 15 mL polyethylene tube and were diluted 1:25 with a diluent solution consisting of 0.1% (v/v) Triton X-100, 0.1% (v/v) HNO_3,2mg/L AU, and internal standards (20 µg/L). The mixture was sonicated for 10min before ICP-MS analysis. Multi-element determination was performed on an Agilent 7700x ICP-MS (Agilent Technologies, Tokyo, Japan) equipped with an octupole reaction system (ORS) collision/reaction cell technology to minimize spectral interferences. The continuous sample introduction system consisted of an autosampler, a quartz torch with a 2.5-mmdiameter injector with a Shield Torch system, a Scott double-pass spray chamber and nickel cones (Agilent Technologies, Tokyo, Japan). A glass concentric MicroMist nebuliser (Agilent Technologies, Tokyo, Japan) was used for the analysis of diluted samples.

Water-soluble Vitamins Measurements

200 µl plasma was deproteinized with 600 µl methanol (Merck), water, acetic acid (9:1:0.1) containing internal standards, thiamine-(4-methyl-13C-thiazol-5-yl-13C3) hydrochloride (Sigma-Aldrich), levomefolic acid-13C, d3, riboflavin-13C,15N2, 4-pyridoxic acid-d3 and pantothenic acid-13C3,15N hemi calcium salt (Toronto Research Chemicals). 500 µl supernatant were dried by nitrogen flow. 60 µl water were added to

the residues, vortexed, the mixture was centrifuged and the supernatant was for analysis. The analysis was performed by UPLC coupled to a Waters Xevo TQ-S Triple Quad mass spectrometry (Waters, USA) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.7 μm, 2.1 × 50 mm) was used for water-soluble vitamins separation with a flow rate at 0.45 ml/min and column temperature of 45 °C. The mobile phases were (A) 0.1 % formic acid in water, (B) 0.1% formic acid in methanol. The following elution gradient was used: 0–1 min,99.0%–99.0% A; 1–1.5 min, 99.0% A–97.0% A; 1.5–2 min, 97.0% A–70.0% A,2–3.5 min, 70% A–30% A; 3.5–4.0 min, 30%A–10.0%A; 4.0–4.8 min, 10%A–10.0%A; 4.9–6.0 min, 99.0%A–99.0%A. Multiple Reaction Monitoring (MRM) was used to monitor all water-soluble vitamins. The mass parameters were as follows, the capillary voltages of 3000V and source temperature of 150°C were adopted. The desolvation temperature was 500°C. The collision gas flow was set at 0.10 ml/min. The cone gas and desolvation gas flow were 150 l/h and 1000 l/h, respectively. All water-soluble vitamins standards were purchased from Sigma-Aldrich (USA).

Fat-soluble Vitamins Measurements

250 µl plasma was deproteinized with 1000 µl methanol and acetonitrile, (v/v,1:1) (Fisher Chemical) containing internal standards, all-trans-Retinol-d5, 25-HydroxyVitamin-D2-d6, 25-HydroxyVitamin-D3-d6, vitamin K1-d7, α -Tocopherol-d6 (Toronto Research Chemicals). 900 µl supernatant was dried by nitrogen flow. 80 µl 80% acetonitrile added to the residues, vortexed, the mixture was centrifuged and the supernatant was for analysis. The analysis was performed by UPLC coupled to an AB Sciex Qtrap 4500 mass spectrometry (AB Sciex, USA) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Waters ACQUITY UPLC BEH C18 column (1.7 µm, 2.1 × 50 mm) was used for fat-soluble vitamins separation with a flow rate at 0.50 ml/min and column temperature of 45 °C. The mobile phases were (A) 0.1 % formic acid in water, (B) 0.1% formic acid in acetonitrile. The following elution gradient was used: 0–0.5 min,60.0%–60.0% A; 0.5–1.5 min, 60.0% A–20.0% A; 1.5–2.5 min, 20.0% A–0% A,2.5–4.5 min, 0% A–0% A; 4.5–4.6 min, 0%A–60.0%A; 4.6–5.0 min, 60.0%A–60.0%A. Multiple Reaction Monitoring (MRM) was used to monitor all fat-soluble vitamins. The mass parameters were as follows, Curtain gas flow 30 l/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 40 l/min, Ion Source Gas 2 (GS 2) flow 50 l/min, Nebulizer Current (NC) 5, temperature 400 °C. All fat-soluble vitamins standards were purchased from Sigma-Aldrich (USA), Toronto research chemical (TRC).

Immune indices Measurements

10 ml whole blood was centrifuged at 3,000 r/min for 10 min, then 165 μl buffy coat was obtained to extract DNA using MagPure Buffy Coat DNA Midi KF Kit (Magen, China). The DNA was then sequenced on BGISEQ- 500 platform in 200 bp singled-end reads. The data processing was using Immune IMonitor[20].

Medical Parameters

All the volunteers were recruited during the physical examination. The medical test including blood tests, urinalysis, routine examination of cervical secretion and so on. All the medical parameters were measured by the physical examination center and shown in **Sup. Data 1**.

Facial Skin feature

The volunteers were required to clean their face without makeup after got up in the morning. The volunteer 's frontal face was photographed by VISIA-CRTM imaging system (Canfield Scientific, Fairfield, NJ, USA) which is equipped with chin supports and forehead clamps that fix the face during the photographing process and maintain a fixed distance between the volunteers and the camera at all times. Eight indices were obtained including spots, pores, wrinkles, texture, UV spots, porphyrins, brown spots and red area from the cheek and forehead, respectively (**Sup. Data 1**). The percentile of index was calculated based on the index value ranked in the age-matched database (**Sup. Data 1**). The higher the percentile of index, the better the facial skin appears.

Physical fitness test

8 kinds of tests were performed to evaluate volunteers' physical fitness condition (**Sup. Data 1**). Vital capacity was measured by HK6800-FH (Hengkangjiaye, China). Eye-closed and single-legged standing was measured by HK6800-ZL. Choice reaction time was measured by HK6800-FY. Grip strength was measured by HK6800-WL. Sit and reach was measured by HK6800-TQ. Sit-ups was measured by HK6800-YW. Step index was measured by HK6800-TJ. Vertical jump was measured by HK6800-ZT. We got a measure value from each test. Then each measure value score was assigned grades A through E based on its corresponding age-matched database.

STATISTICAL ANALYSIS

Quality control, Taxonomic annotation and abundance calculation

The sequencing reads were quality-controlled as described previously[19]. The more stringent condition for removal of host sequences was used for cervical samples[3], through alignment to the GRCh38 reference.

Taxonomic assignment of the high-quality cervical metagenomic data was performed using MetaPhlAn2[21].

Taxonomic assignment of the high-quality faecal metagenomic data was performed using the reference gene catalog comprising 9,879,896 genes[22]. Taxonomy of the vagino-cervical MLGs and the faecal MGSs were then determined from their constituent genes, as previously described[18,23–25].

Random-forest on the influence of female life history factors

The factors in female life history questionnaire were fitted against the relative abundances of metaphlan2 profile (found in at least 10% of the samples) in the cervical samples using default parameters in the RFCV regression function from randomForest package in R. All the categorical variables were converted into continue variables, and nominal variables were converted into dummy variables. Spearman's correlation between measured value and 5-fold cross-validation predicted value was calculated, then rank the key predictable factors.

The global effect size between vagino-cervical microbiome and omics data.

To evaluate the combined effect size of vagino-cervical microbiome on omics data, we used forward stepwise redundancy analysis of omics data lists on the relative abundances of metaphlan2 profile in forward.sel function in the packfor package in R. This analysis provided a global versus global association between any two omics datasets that maximize the associations by use the most predict power non-redundant predictors.

The factors in each type of omics predicted by vagino-cervical microbiome

The factors in each type of omics were fitted against the relative abundances of metaphlan2 profile (found in at least 10% of the samples) in the cervical samples using default parameters in the RFCV regression function from randomForest package in R. All the categorical variables were converted into continue variables, and nominal

variables were converted into dummy variables. Spearman's correlation between measured value and 5-fold cross-validation predicted value was calculated, then rank the top 8 predictable factors in each type.

Wisdom of crowds for robust network construction between vagino-cervical microbial species and other omics data

A new Multi-omics analyses method[10] was used to integrated coefficient of linear regression, variance importance from randomForest and Spearman's correlation to construct omics flux networks and then visualized in CytoScape. The detail as indicated: Step 1: Data processing. All categorical variables in other omics data were converted into continuous variables, and nominal variables were converted into dummy variables. Missing values were filled with median, the samples which contained more than 70% missing variables were removed. The microbial species less than10% in all the samples were also removed. Removed near zero variable variables. For linear models, variables were normalized. Outliers were defined as outside of the 95% quartiles and removed. Step 2: Method implementation. Random forest variable importance was used to identify the most important predictor variables[26]. RFCV regression function from randomForest package in R with default parameter was used to get the 5-fold average variable importance. We calculated the Spearman's correlation with the cor.test function in base R software. For linear regression, we considered penalty regression to overcome the sparse and co-linear problem, cv.glmnet function from glmnet package in R was first used to figure out the best lambda parameter, bootstrapping glmnet with 0.632 re-sampling was performed 100 times, then we obtained the best lambda.

Step 3: Construction of robust networks. We kept first 5 average ranks for each target variable and retained edges with Spearman's correlation Q-value <0.1. Then ggplot package in R was used to make barplot for some representative female life history factors (**Fig 3**). CytoScape was also used to visualized the omics network (**Extended Data Fig. 4**).

Association between microbiome pathways and other omics

Pathway profile was calculated from the vagino-cervical metagenomic data using humann2. Spearman's correlation was calculated between the relative abundance of each pathway and other numerical data collected. R package heatmap were used for visualization. Q-value <0.1 was considered as significant.

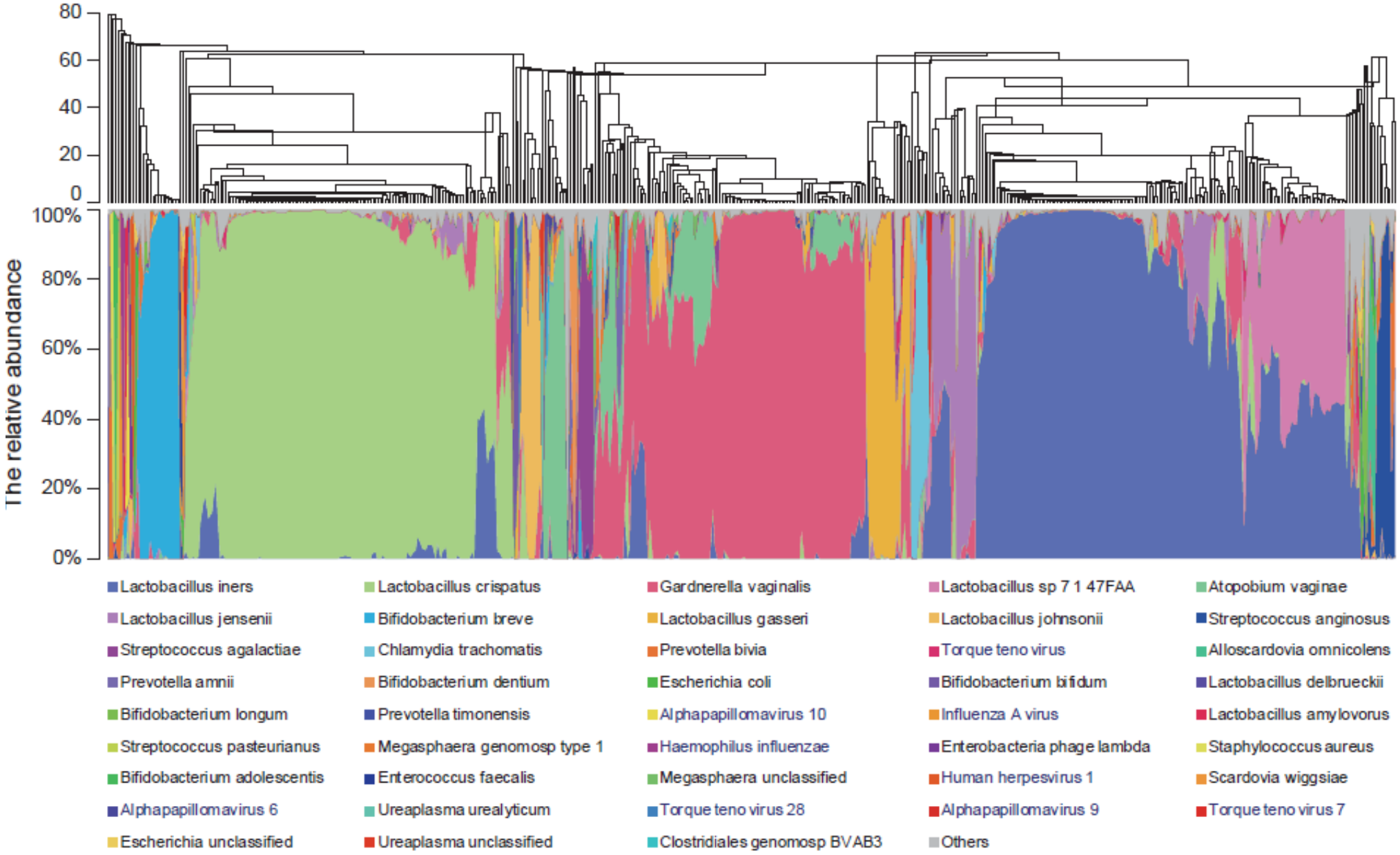

Figure 1 Vagino-cervical microbiome of the cohort.

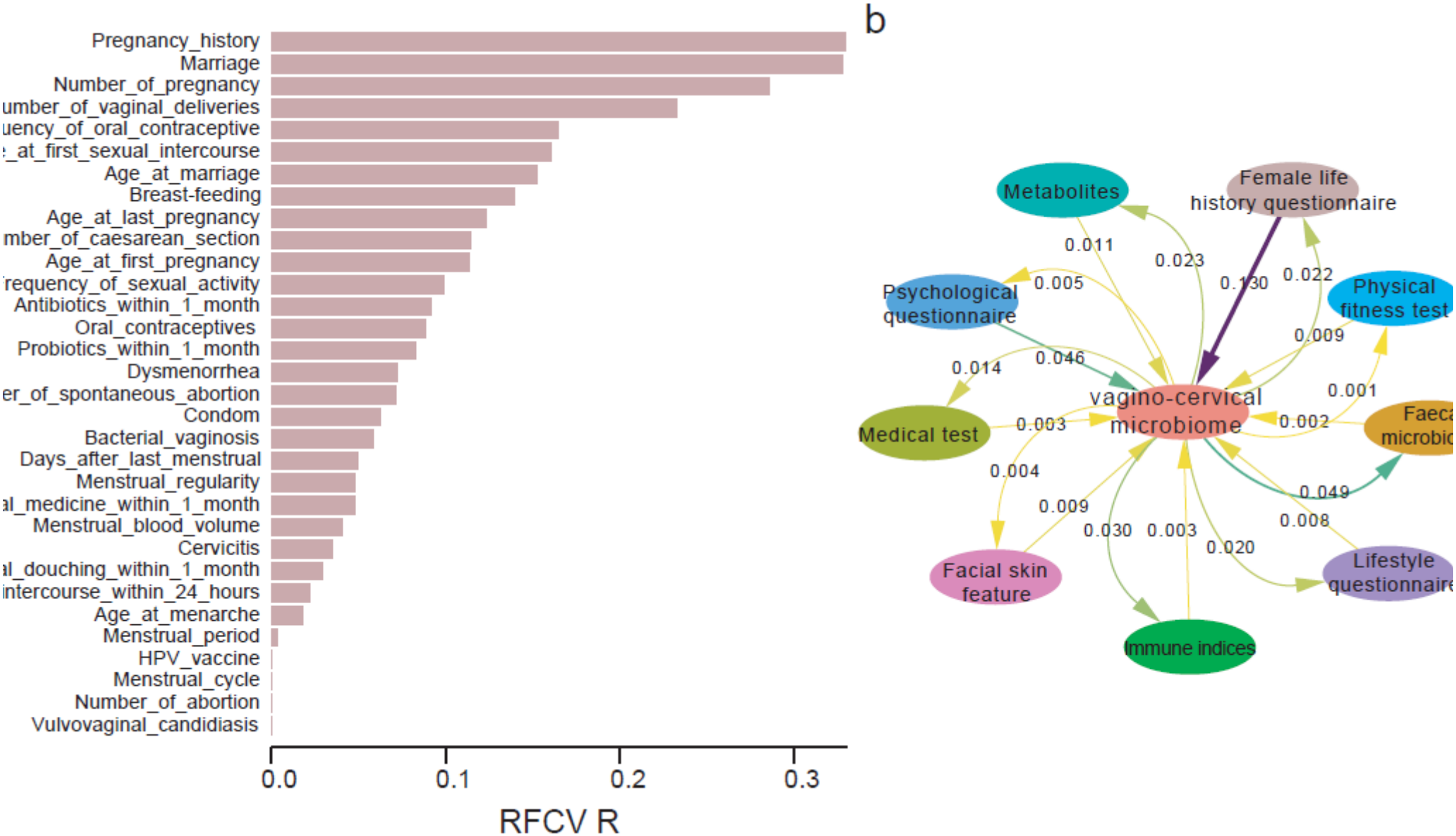

Figure 2 Global view of factors influencing the vagino-cervical microbiome.

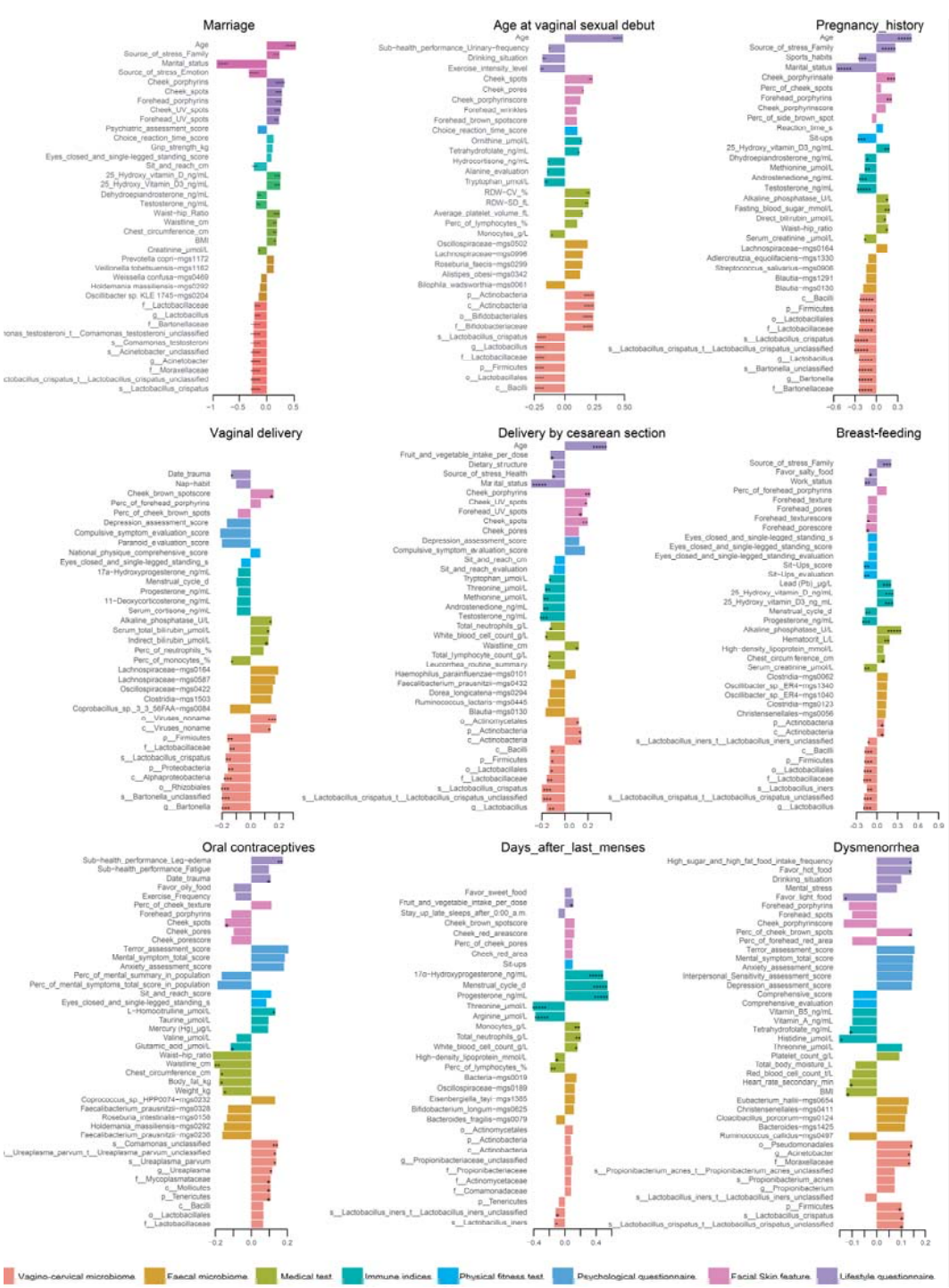

Figure 3 Specific influences from reproductive factors on omics data.