



## ORIGINAL RESEARCH

# Life History Recorded in the Vagino-cervical Microbiome Along with Multi-omes



Zhuye Jie<sup>1,2,3,#</sup>, Chen Chen<sup>1,2,3,\*,#</sup>, Lilan Hao<sup>1</sup>, Fei Li<sup>1</sup>, Liju Song<sup>1</sup>, Xiaowei Zhang<sup>1</sup>, Jie Zhu<sup>1</sup>, Liu Tian<sup>1</sup>, Xin Tong<sup>1</sup>, Kaiye Cai<sup>1,4</sup>, Zhe Zhang<sup>1</sup>, Yanmei Ju<sup>1,5</sup>, Xinlei Yu<sup>1</sup>, Ying Li<sup>1</sup>, Hongcheng Zhou<sup>6</sup>, Haorong Lu<sup>6</sup>, Xuemei Qiu<sup>1</sup>, Qiang Li<sup>1</sup>, Yunli Liao<sup>1</sup>, Dongsheng Zhou<sup>1</sup>, Heng Lian<sup>1</sup>, Yong Zuo<sup>1</sup>, Xiaomin Chen<sup>1</sup>, Weiqiao Rao<sup>1</sup>, Yan Ren<sup>1</sup>, Yuan Wang<sup>1</sup>, Jin Zi<sup>1</sup>, Rong Wang<sup>1</sup>, Na Liu<sup>1</sup>, Jinghua Wu<sup>1</sup>, Wei Zhang<sup>1</sup>, Xiao Liu<sup>1</sup>, Yang Zong<sup>1</sup>, Weibin Liu<sup>1</sup>, Liang Xiao<sup>1,4,7</sup>, Yong Hou<sup>1</sup>, Xun Xu<sup>1</sup>, Huanming Yang<sup>1,8</sup>, Jian Wang<sup>1,8</sup>, Karsten Kristiansen<sup>1,3,7</sup>, Huijue Jia<sup>1,2,\*</sup>

<sup>1</sup> BGI-Shenzhen, Shenzhen 518083, China

<sup>2</sup> Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, BGI-Shenzhen, Shenzhen 518083, China

<sup>3</sup> Department of Biology, University of Copenhagen, Copenhagen DK-2100, Denmark

<sup>4</sup> Shenzhen Engineering Laboratory of Detection and Intervention of Human Intestinal Microbiome, BGI-Shenzhen, Shenzhen 518083, China

<sup>5</sup> BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China

<sup>6</sup> China National Genebank, BGI-Shenzhen, Shenzhen 518120, China

<sup>7</sup> BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

<sup>8</sup> James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

Received 21 December 2020; accepted 31 January 2021

Available online 9 June 2021

Handled by Jun Yu

## KEYWORDS

Vagino-cervical microbiome;  
Metagenomic shotgun  
sequencing;  
Pregnancy history;

**Abstract** The vagina contains at least a billion microbial cells, dominated by lactobacilli. Here we perform **metagenomic shotgun sequencing** on cervical and fecal samples from a cohort of 516 Chinese women of reproductive age, as well as cervical, fecal, and salivary samples from a second cohort of 632 women. Factors such as **pregnancy history**, **delivery history**, cesarean section, and **breastfeeding** were all more important than menstrual cycle in shaping the microbiome, and such

\* Corresponding authors.

E-mail: jiahuijue@genomics.cn (Jia H), chenchen1@genomics.cn (Chen C).

# Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2021.01.005>

1672-0229 © 2022 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Delivery history;  
Breastfeeding

information would be necessary before trying to interpret differences between **vagino-cervical microbiome** data. Greater proportion of *Bifidobacterium breve* was seen with older age at sexual debut. The relative abundance of lactobacilli especially *Lactobacillus crispatus* was negatively associated with pregnancy history. Potential markers for lack of menstrual regularity, heavy flow, dysmenorrhea, and contraceptives were also identified. Lactobacilli were rare during breastfeeding or post-menopause. Other features such as mood fluctuations and facial speckles could potentially be predicted from the vagino-cervical microbiome. Gut and salivary microbiomes, plasma vitamins, metals, amino acids, and hormones showed associations with the vagino-cervical microbiome. Our results offer an unprecedented glimpse into the microbiota of the female reproductive tract and call for international collaborations to better understand its long-term health impact other than in the settings of infection or pre-term birth.

## Introduction

The human body is a supra-organism containing tens of trillions of microbial cells [1]. Studies in human cohorts and animal models have revealed an integral role played by the gut microbiota in metabolic and immunological functions [2]. Disease markers as well as prebiotic or probiotic interventions are being actively developed [3,4]. The microbiota studies on other mucosal sites such as the vagina and the mouth also have great potential for human health [5–7]. Despite continued controversy, the presence of microorganisms beyond the cervix (*i.e.*, the upper reproductive tract) is increasingly recognized even in non-infectious conditions [8–12], with much debated implications for women's and infants' health.

Lactobacilli have long been regarded as the keystone species of the vaginal microbiota [13]. Lactic acid produced by these microorganisms helps maintain a low vaginal pH of 3.5–4.5, and wards off pathogenic microorganisms [14]. Prevention of the human immunodeficiency virus (HIV) infection and other sexually transmitted infections (STIs), preterm birth, and bacterial vaginosis (BV) have been major efforts [14–16]. Germ-free mice treated with *Lactobacillus crispatus* had fewer activated CD4<sup>+</sup> T cells in the genital tract compared to those treated with *Prevotella bivia*, explaining the difference in HIV acquisition in human cohorts [17]. How vaginal lactobacilli interplay with *Candida albicans* and other related fungi is also an important question for preventing or treating vulvovaginal candidiasis [18]. The fact that human sequences make up more than 90% of the reads in female reproductive tract samples, in contrast to 1% in feces [4,9,19], has made metagenomic shotgun sequencing of reproductive tract samples more expensive. To our knowledge, all studies of the vaginal microbiota other than the Human Microbiome Project (HMP) used 16S rRNA gene amplicon sequencing or quantitative polymerase chain reaction (qPCR) [6,14,19,20], which lacked a view of the overall microorganism community including bacteria, archaea, viruses, and fungi [21], as well as the encoded functional capacity. Besides infection, studies on the vaginal microbiota in relation to the current sexual activity or the menstrual cycle have also been reported [13,22]. However, lasting impacts from other potentially important factors such as sexual debut, pregnancy, and breastfeeding have not been investigated in a large cohort.

As a reservoir of microbes instead of a transient entity [22], the female vagino-cervical microbiota might also reflect conditions in other body sites. It is however not clear whether metabolites and cells in circulation or microbes in other mucosal sites cross-talk with the vagino-cervical microbiome.

Intersecting with hormones, metabolic functions, and immune functions, we find it intriguing to explore the potential link of the vagino-cervical microbiome to the brain and the face using questionnaires and measured data that are available for large cohorts.

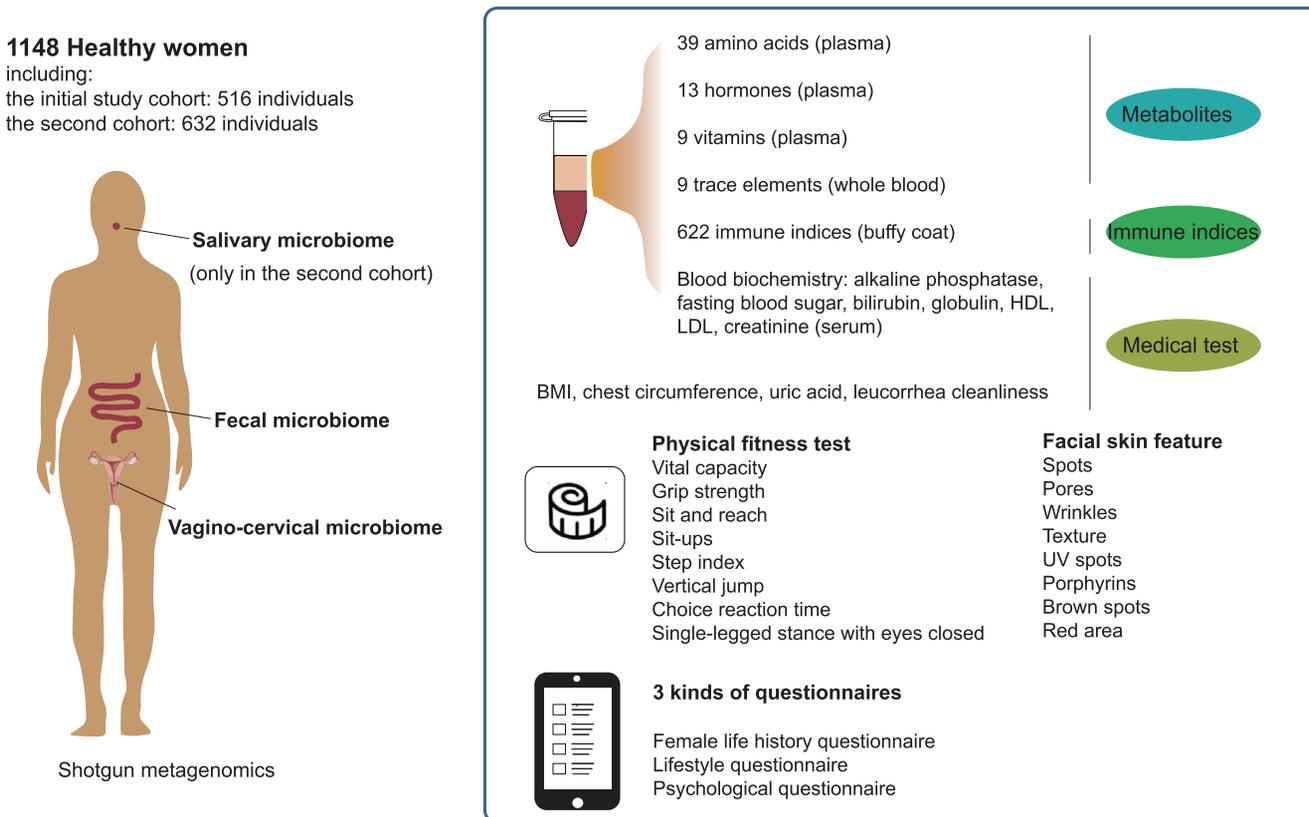
Here, we report the full spectrum of the vagino-cervical microbiome from 1148 healthy women using metagenomic shotgun sequencing. As part of the 4D-SZ cohort (trans-omics, with more time points in future studies, based in China), we also comprehensively measured the parameters including fecal/salivary microbiomes, plasma metabolites, medical test data, immune indices, physical fitness test, and facial skin imaging, as well as female life history questionnaire, lifestyle questionnaire, and psychological questionnaire (**Figure 1**). Our work pinpoints other metadata or multi-omes that can predict or be predicted from the microbiota in the female reproductive tract, which would illuminate future designs of population cohorts, mechanistic investigations, and means of intervention.

## Results

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

To explore the vagino-cervical microbiome, 516 healthy Chinese women were recruited during a physical examination as the initial study cohort [median age 30, 95% confidence interval (CI): 21–39] (Figure 1; Table S1). Metagenomic shotgun sequencing was performed on the cervical samples, and high-quality non-human reads were used for taxonomic profiling of the vagino-cervical microbiome (**Figure 2A**; Table S2).

In agreement with the 16S rRNA gene amplicon sequencing data from the USA [13], the vagino-cervical microbiota of this Asian cohort was lactobacilli-dominated. In our study, the species whose relative abundance accounted for 50% or higher in an individual was defined as a community type, while all species accounting for less than 50% of the microbiota in an individual were collectively identified as a diverse community type. The community types characterized by *L. crispatus* and *Lactobacillus iners*, accounting for 24.61% and 22.67% of the individuals, respectively, were the two of the most common types in our initial study cohort (Figure 2A, Figure S1A). *L. iners*, which although traditionally viewed as 'healthier' than non-lactobacilli species such as *Gardnerella vaginalis*, has been shown to confer far less protection against bacterial and viral infections than *L. crispatus* [17,23]. *G. vaginalis* could be



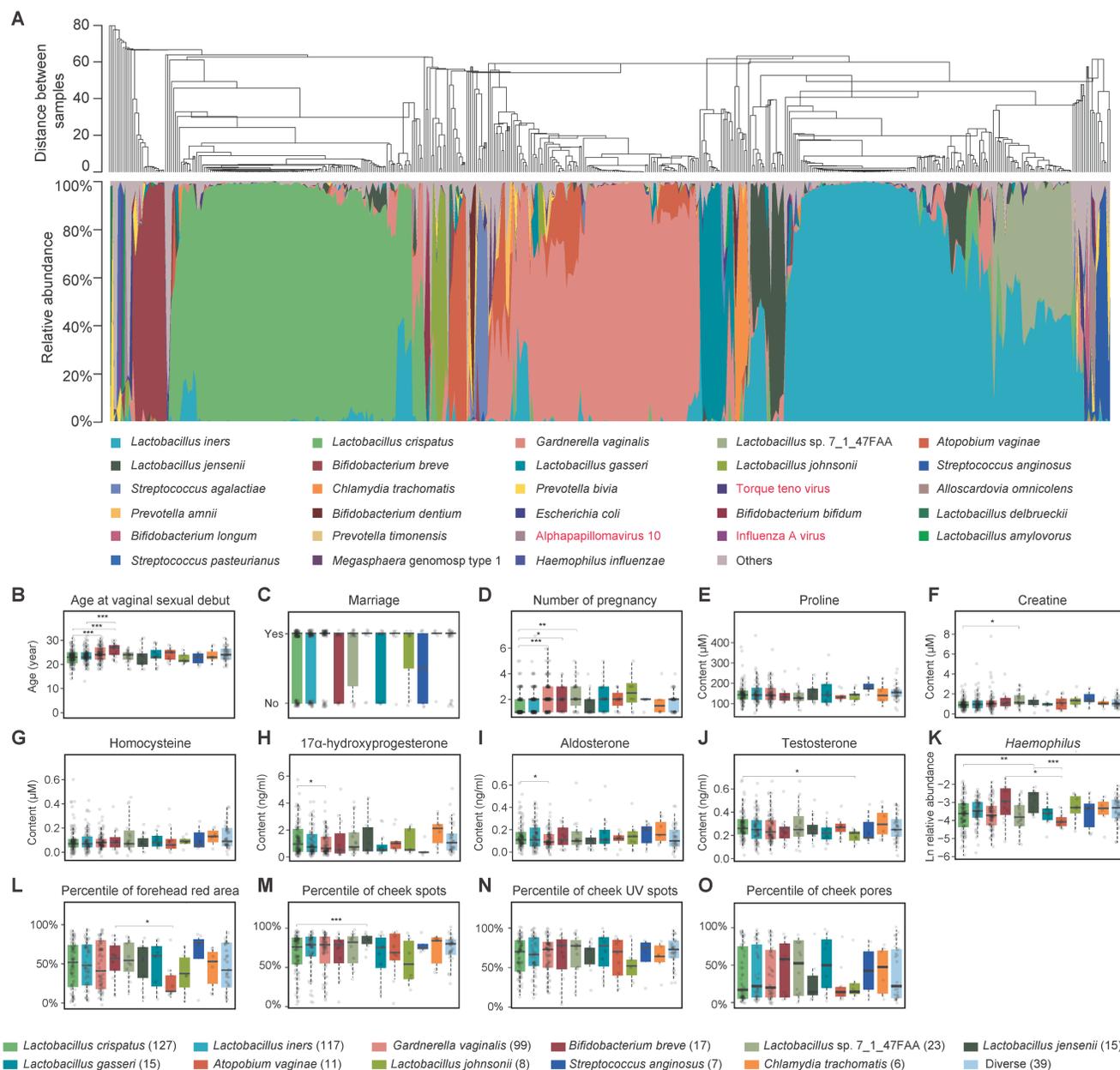
**Figure 1 Schematic overview of the study**

The study consists of two cohorts with varied characteristics. The 516 volunteers recruited in 2017 were regarded as the initial study cohort. The 632 volunteers recruited in 2018 were regarded as the second cohort. Vagino-cervical microbiome, fecal microbiome, metabolites, medical test data, physical fitness test data, facial skin features, and three kinds of questionnaires were collected in both cohorts. Salivary microbiome was only available in the second cohort; among medical test results, InBody results, which measure body composition, were only available in the initial study cohort; the woman life history questionnaires differ by a few terms (Table S1). HDL, high density lipoprotein; LDL, low density lipoprotein; BMI, body mass index; UV, ultraviolet.

detected in 63.57% of the individuals, and in 19.19% of the 516 women, the relative abundance of *G. vaginalis* was equal to or higher than 50% (Figure 2A, Figure S1A; Table S3). *Atopobium vaginae* (recently renamed as *Fannyhessaea vaginae* [24]) and *G. vaginalis* are commonly considered to co-occur in BV [20,25]. This co-occurrence was also observed here, *i.e.*, 90.16% of the volunteers with *A. vaginae* in their cervical samples harbored *G. vaginalis* (Figure 2A, Fisher's exact test,  $P = 1.237E-13$ , odds ratio = 7.37603). Yet, there were individuals who were dominated by *A. vaginae* (Figure 2A, Figure S1A). Rare subtypes (an identified community type in less than 5% of the 516 individuals) such as those characterized by *Bifidobacterium breve* (3.29%), *Lactobacillus jensenii* (2.91%), *Lactobacillus gasseri* (2.91%), *A. vaginae* (2.13%), *Lactobacillus johnsonii* (1.55%), *Streptococcus anginosus* (1.36%), and *Chlamydia trachomatis* (1.16%) were also detected in this cohort (Figure 2A, Figure S1A). *Streptococcus agalactiae* (Group B *Streptococcus*), a bacterium responsible for neonatal sepsis and recently reported in placenta [10], could be detected in 5.62% of the individuals (Figure 2A; Table S3). Fungal species including *C. albicans*, *Candida glabrata*, *Candida dubliniensis*, and *Candida tropicalis*, commonly believed to result in vulvovaginal candidiasis [26], were also detected in 3.1% of the individuals. Other microorganisms including *P. bivia*, *Escherichia coli*, *Ureaplasma parvum*, human

papillomavirus (HPV), herpesviruses, Influenza A virus, and *Haemophilus influenzae* were abundant in some individuals (Figure 2A; Table S3).

Individuals with a vagino-cervical microbiome dominated by different bacteria showed significant differences in some of the questionnaire or omic data [generalized linear model (GLM) likelihood ratio test, overall  $P$  value < 0.05; Student's  $t$ -test between two groups; Figure 2B–O, Figure S1A]. The community type characterized by *L. crispatus* was overrepresented in women who had fewer pregnancies than women with the community type *G. vaginalis*, *B. breve*, or *Lactobacillus* sp. 7\_1\_47FAA (Figure 2D), while individuals of the *B. breve* type had older age at vaginal sexual debut compared to those with the community type *L. crispatus* or *L. iners* (Figure 2B). The plasma concentrations of  $17\alpha$ -hydroxyprogesterone and aldosterone were higher in individuals of the *L. crispatus* type than in the *G. vaginalis* type (Figure 2H and I), while the plasma concentration of testosterone was higher in individuals of the *L. crispatus* type than in the *L. johnsonii* type (Figure 2J). The relative abundance of fecal *Haemophilus* spp. was higher in individuals of the *L. jensenii* type than in the *L. crispatus* and *A. vaginae* community types (Figure 2K). Women of the *B. breve* type had a more serious facial skin problem of red area on the forehead than women with the *A. vaginae* type (Figure 2L), while women with the *L. crispatus* type had fewer



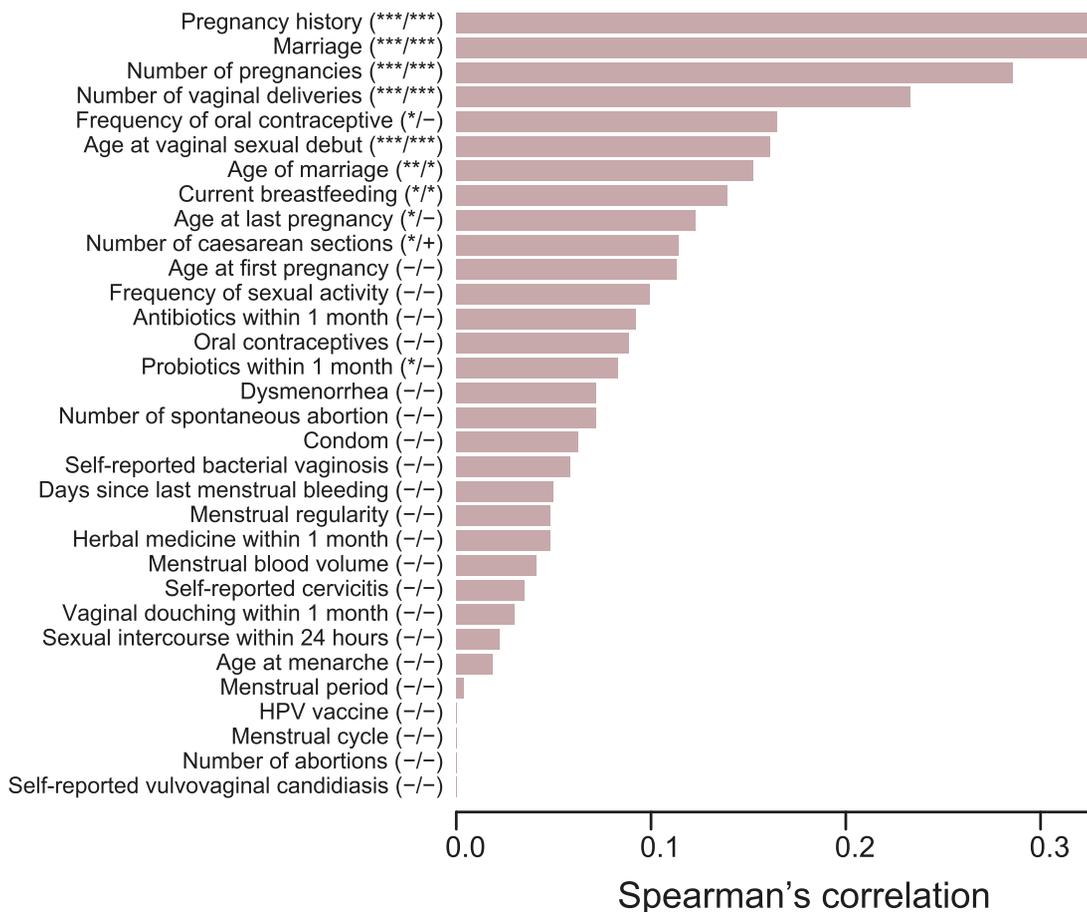
**Figure 2 Vagino-cervical microbiome of the initial study cohort**

**A.** The microbial composition in each sample at the species level according to MetaPhlan2 is shown. The dendrogram in (top) was a result of a centroid lineage hierarchical clustering based on Euclidean distances between the microbial composition proportion. Black and red taxa labels denote bacteria and viruses, respectively. **B–O.** Specific multi-omic factors showing significant differences among 12 key vagino-cervical microbiota types estimated by GLM likelihood ratio test ( $P < 0.05$ , 554 comparisons). Colors of the bars in (B–O) denote the microbiota types as listed at the bottom of the figure. The number of samples per microbiota type is labeled in the parentheses. The horizontal lines with  $P$  values mark the pairwise post hoc comparisons using Student’s  $t$ -test (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ). The genus in (K) belongs to fecal microbiome. Boxes denote the IQR between the first and third quartiles (25th and 75th percentiles, respectively), and the line inside the boxes denotes the median. The whiskers denote the lowest and highest values within 1.5 times the IQR from the first and third quartiles, respectively. GLM, generalized linear model; IQR, interquartile range.

spots on the cheeks than women with the *L. jensenii* type (Figure 2M).

According to the metagenomic shotgun data, the mean proportion of non-bacterial sequences was 3.45% (Figure 2A). The vagino-cervical community types of *G. vaginalis*, *L. crispatus*, and *B. breve* showed lower proportions of human

sequences than other community types (e.g.,  $P = 3.9E-6$  between *L. crispatus* and *L. iners* community types,  $P = 1.7E-8$  between *G. vaginalis* and *L. iners* community types, and  $P = 0.0036$  between *B. breve* and *L. gasseri* community types, Wilcoxon ranked sum test, Figure S1B), suggesting that in future studies lower amounts of sequencing could be



**Figure 3** Factors from the female life history questionnaire influencing the vagino-cervical microbiome in the initial study cohort

Female life history questionnaire entries on the vagino-cervical microbiome, ordered according to their respective 5-fold cross-validated random forest importance on the microbiome composition. X-axis (length of the bar) is the model performance measured as the Spearman's correlation coefficient between the prediction and measurement. In the parentheses after Y-axis labels, the first column symbols are  $P$  values with 999 permutations, and the second column symbols are  $Q$  values for BH method (32 comparisons). “-” denotes value  $\geq 0.1$ , “+” denotes  $0.05 \leq \text{value} < 0.1$ , “\*” denotes value  $< 0.05$ , “\*\*\*” denotes value  $< 0.01$ , and “\*\*\*\*” denotes value  $< 0.001$ . BH, Benjamini-Hochberg; HPV, human papillomavirus.

used for these compared to other community types, and that species like *L. iners* and *L. gasseri* might be more engaged with human cells.

#### Factors associated with the vagino-cervical microbiome

We computed the prediction value (5-fold cross-validated random forest model) of each factor independently in the female life history questionnaire on the relative abundance of vagino-cervical microbiome data, and found the most important factors to be pregnancy history, marriage, number of pregnancies, number of vaginal deliveries, and age at vaginal sexual debut, followed by age of marriage, and current breastfeeding [ $P < 0.05$  with 999 permutations and  $Q < 0.05$  for Benjamini-Hochberg (BH) method] (Figure 3, Figure S2). None of these factors significantly correlated with one another (Figure S3; Table S4).

These important factors for the vagino-cervical microbiome were validated in an independent cohort of 632 individuals that differed in age distribution as well as sequencing mode (median age 32, 95% CI: 18–47; paired-end 100 bp instead

of single-end 50 bp for vagino-cervical samples) (Figure 1, Figure S3; Tables S1–S3). Questionnaire entries such as pregnancy history, marital status, current breastfeeding, and mode of the most recent delivery (caesarean or vaginal) were again found as the most important factors to exhibit correlations with the vagino-cervical microbiome (Figure S3). Duration of current breastfeeding emerged as a strong predictor of the vagino-cervical microbiome composition, which reflected the differences in questionnaire design that was only available in the second but not the first cohort (Figures S3 and S4; Table S1). The prediction value of menstrual cycle was slightly augmented with the presence of postmenopausal women in the second cohort (Figure 3, Figure S3).

Forty-five of the volunteers in the second cohort were postmenopausal (median age 54, 95% CI: 44–64), an age group untouched by HMP. Metagenomic shotgun data revealed diminished lactobacilli in their vagino-cervical microbiome, while the mean proportion of viral sequences reached 37% (Figure S5). HPV was not the most abundant or prevalent virus in these postmenopausal individuals; herpes simplex virus (HSV), phages, and torque teno virus (TTV) were also

detected (Figure S5). The endogenous retrovirus was likely part of the human genome, which would need further validation (Figure S5). In contrast, there was no sign of more *C. albicans* or other fungi, suggesting that the lack of glycogen in postmenopausal individuals might also have counteracted fungal growth (Figure S5). Three of the individuals were dominated by *L. crispatus* and two of the individuals by *L. iners*, none of whom was reported to receive hormone replacement therapy, implying genuine individual differences (Figure S5C). The non-lactobacilli bacterial species are also known as vaginal or salivary species, with overgrowth of *E. coli* in two of the individuals (Figure S5C). The taxonomic profile remained robust when we arbitrarily trimmed the paired-end 100 bp data to single-end 50 bp or single-end 100 bp (Spearman's correlation coefficient = 0.998 between Pseudo-SE50 and Pseudo-SE100, 0.993 between Pseudo-SE50 and PE100) (Figure S6).

As a whole, the vagino-cervical microbial composition showed the greatest explained variances for these questionnaire data collected for the vagino-cervical samples, followed by other data collected on the same day, such as fecal microbiome composition, psychological questionnaire, plasma metabolites, immune indices, facial skin imaging, medical test data, and physical fitness test results (Figure 4A). Age, pregnancy history, marriage, and number of pregnancies were the most significant factors among all the multi-omic data [with the largest Spearman's correlation coefficients between random forest cross validation (RFCV) prediction and observed data,  $P < 0.001$ , and  $Q < 0.001$ ; Figure 4B]. *L. crispatus* in the fecal microbiome was most predictive of the vagino-cervical microbiome ( $P < 0.001$  and  $Q < 0.001$ ); plasma phosphoserine ( $P < 0.001$  and  $Q < 0.001$ ), L-homocitrulline ( $P < 0.01$  and  $0.05 \leq Q < 0.1$ ), and testosterone ( $P < 0.01$  and  $0.05 \leq Q < 0.1$ ) were ranked at the top among metabolites (Figure 4B); weaker signals were observed for serum albumin and low density lipoprotein (LDL) in medical test data ( $P < 0.01$  but  $Q \geq 0.1$ ) and spots on the cheeks in facial skin imaging ( $P < 0.05$  but  $Q \geq 0.1$ ) (Figure 4B).

### Specific influences from pregnancy history, contraception, and menstrual symptoms

Marital status was one of the most significant factors to associate with the vagino-cervical microbiome (Figure 3, Figure S3). The married individuals showed negative correlations with relative abundances of *L. crispatus*, *Comamonas testosteroni*, and *Acinetobacter* spp. (Spearman's correlation,  $Q < 0.01$ ) (Figure 5A; Table S6). Compared to unmarried women, married women had higher concentrations of plasma 25-hydroxy vitamin D, and their plasma testosterone, dehydroepiandrosterone (DHEA), and creatinine were lower (Figure 5A; Table S6). *C. testosteroni* TA441, which at the species level was correlated with being married, has been reported to degrade testosterone [27]. The age at vaginal sexual debut showed positive correlations with Bifidobacteriaceae (which includes *G. vaginalis* and *B. breve*) (Figure 5B; Table S4), and negative associations with *L. crispatus* (Figure 5B; Tables S5 and S6), consistent with analyses based on dominant bacteria (Figure 2B).

Similarly, the women who went through pregnancy showed relatively less *Lactobacillus*, especially *L. crispatus* compared

to nullipara, as well as less *Bartonella* in the vagino-cervical microbiome (Figure 5C; Table S6). We observed increased concentrations of plasma vitamin D3 in women with previous pregnancy, but lower concentrations of plasma testosterone, androstenedione, dehydroepiandrosterone, and methionine (Figure 5C). Vaginal deliveries were associated with decreased *L. crispatus* and *Bartonella* species (Figure 5D; Table S6); cesarean section deliveries were also correlated with less *L. crispatus*, but higher abundance of Actinobacteria (Figure 5E; Table S6). In addition, cesarean sections were correlated with lower plasma concentrations of testosterone, androstenedione, methionine, threonine, and tryptophan, as well as a larger waistline, lower neutrophil and lymphocyte counts, and abnormal leucorrhoea (Figure 5E). Cesarean section deliveries were also associated with lower facial skin score ranking such as ultraviolet (UV) spots and porphyrins, while those with vaginal deliveries appeared with less brown spots on the cheeks (Figure 5D and E).

A total of 137 volunteers (45.99% from the initial cohort, 54.01% from the second cohort) happened to be actively breastfeeding (Figure 5F, Figure S4). Those who were within the first two years of delivery often lacked lactobacilli, especially *L. crispatus*, and were dominated by viral sequences or BV-related species such as *Prevotella* spp. and *Atopobium* spp. (Figure S4). Only 6.57% of the individuals harbored  $> 50\%$  *L. crispatus*, which would represent  $> 20\%$  of the cohort for non-pregnant, non-breastfeeding Asia women of reproductive age (Figures 2 and 5F). In addition, 23.36% of the breastfeeding individuals were dominated by *L. iners*, and 18.25% individuals who had  $> 50\%$  *G. vaginalis*, similar to the overall distribution (Figure 2, Figure S4). Alkaline phosphatase, plasma vitamin D, and lead (Pb) were found to be higher in the breastfeeding individuals, combined with slight reduction of progesterone and lack of menses in lactating women (Figure 5F).

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 *L. gasseri* (Figure 6; Tables S5 and S6). Oral contraceptives, although still rare in our cohort, were associated with increased abundances of *L. iners*, *U. parvum*, and *Comamonas* species (Figures 5G and 6; Tables S5 and S6). *U. parvum* is a bacterium commonly isolated from pregnant women [28] and recently reported in the lower respiratory tract of preterm infants [29] and in preterm placenta [10]. *Comamonas* has been implicated in the fecundity in *Caenorhabditis elegans* and identified as a marker for infertility due to endometriosis [8,30]. As noted above for marital status and testosterone (Figure 5A), *Comamonas testosteroni* TA441 is known to degrade testosterone [27], while an unnamed *Comamonas* species was associated with oral contraceptives (progesterone derivatives) here (Figures 5G and 6). Moreover, oral contraceptives exhibited a positive correlation with plasma homocitrulline (Figure 5G). These results would be worth investigating in large longitudinal cohorts.

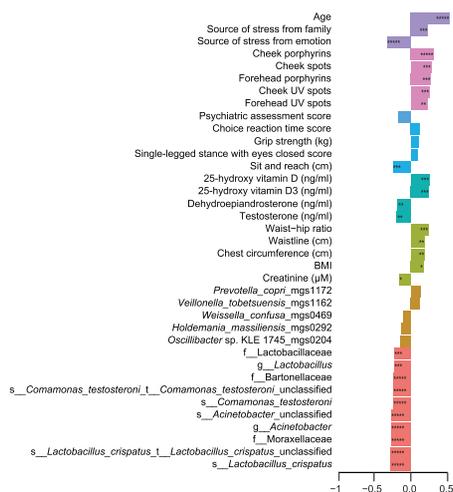
Menstrual phases are known to influence the microbiota in the female reproductive tract [8,22]. We confirmed in our cross-sectional cohort that *L. iners* was relatively more abundant in the proliferative phase (*i.e.*, after menses and before ovulation) than in the secretory phase (*i.e.*, after



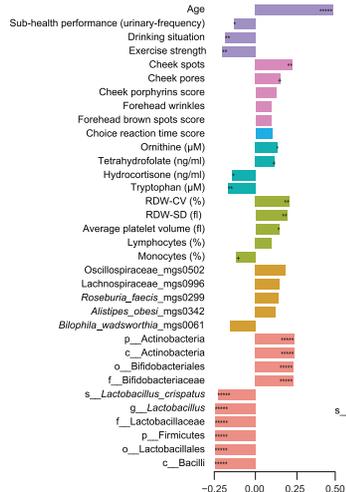
**Figure 4 Global view of factors influencing the vagino-cervical microbiome in the initial study cohort**

**A.** Predicting the vagino-cervical microbiome from each omic dataset and *vice versa* using stepwise redundancy analysis. Numbers on the straight arrows indicate adjusted R-squared from vagino-cervical microbiome to multi-omic data; numbers on the curved arrows indicate adjusted R-squared from multi-omic data to vagino-cervical microbiome. **B.** Top 8 factors in each type of multi-omic data that are predicted by the vagino-cervical microbiome. Each covariate from the multi-omic data is detailed in Table S1. X-axis (length of the bar) is the model performance measured as the Spearman's correlation coefficient between the prediction and measurement. In the parentheses after Y-axis labels, the first column symbols are *P* values with 999 permutations, and the second column symbols are *Q* values for BH method (32 comparisons). “-” denotes value  $\geq 0.1$ , “+” denotes  $0.05 \leq \text{value} < 0.1$ , “\*” denotes value  $< 0.05$ , “\*\*\*” denotes value  $< 0.01$ , and “\*\*\*\*” denotes value  $< 0.001$ .

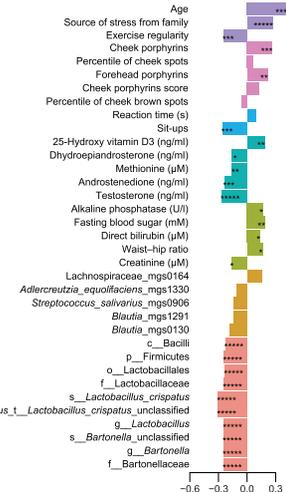
**A Marriage**



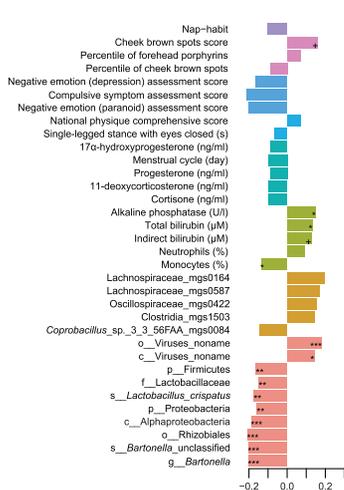
**B Age at vaginal sexual debut**



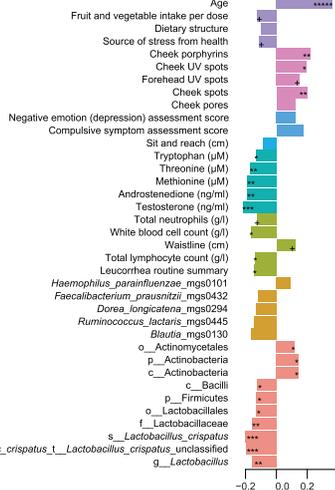
**C Pregnancy history**



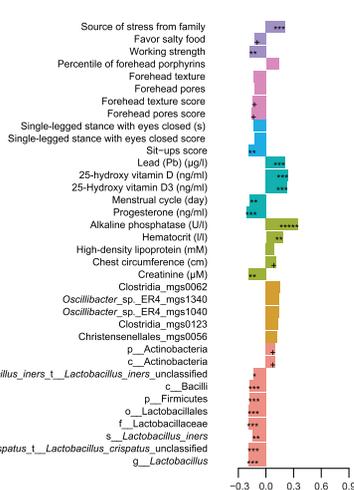
**D Number of vaginal deliveries**



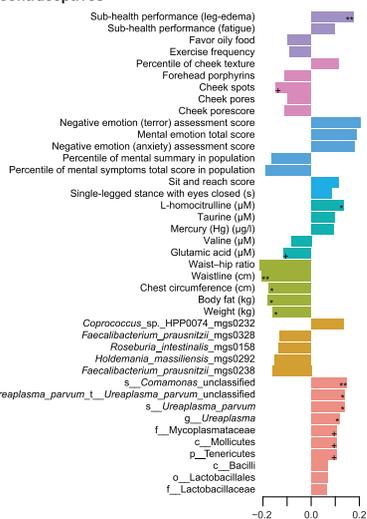
**E Number of cesarean deliveries**



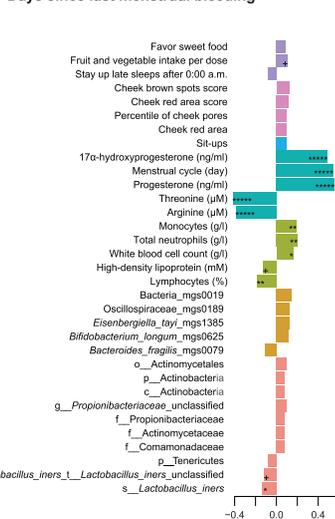
**F Current breastfeeding**



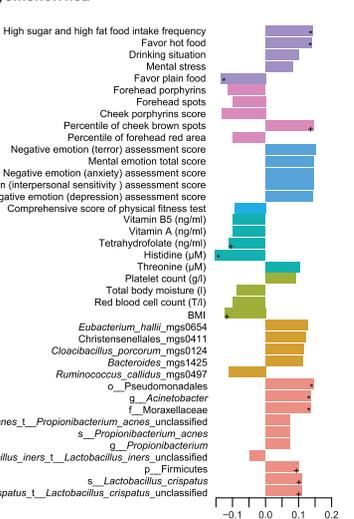
**G Oral contraceptives**



**H Days since last menstrual bleeding**



**I Dysmenorrhea**





**Figure 6 Wisdom of the crowds for the association network between vagino-cervical microbial species and multi-omic data in the initial study cohort**

Results from GLM with penalty (cv.glmnet), random forest cross validation, and Spearman’s correlation are integrated and then visualized in Cytoscape. Red edges indicate negative associations; cyan edges indicate positive associations.

ovulation and before menses, waiting for implantation of embryos), while *L. crispatus* was relatively more abundant in the secretory phase than in the proliferative phase, coinciding with dynamics in progesterone as well as threonine and arginine (Figure 5H, Figure S7). White blood cell (WBC) counts

also recovered after menses (Figure 5H). These results are consistent with the days after menses being more susceptible to BV relapse and a more stable vagino-cervical microbiota during or after ovulation [22]. Individuals with self-reported regular periods showed negative correlations with *Lactobacillus*

**Figure 5 Specific influences of reproductive factors on multi-omic data in the initial study cohort**

The factors shown are marriage (A), age at vaginal sexual debut (B), pregnancy history (C), number of vaginal deliveries (D), number of cesarean deliveries (E), current breastfeeding (F), oral contraceptives (G), days since last menstrual bleeding (H), and dysmenorrhea (I). The bars are colored according to the type of omic data, as shown in Figure 3. The metabolites such as amino acids, hormones, and vitamins were measured in plasma, and trace elements were measured in whole blood. The blood biochemistry such as alkaline phosphatase, fasting blood sugar, direct bilirubin, creatinine, total bilirubin, and HDL was measured in serum. Each covariate from the multi-omic data is detailed in Table S1. The length of the bars represents the Spearman’s correlation coefficient between the respective factor and the multi-omic data. Average rank selects 13,425 edges from 66,301 associations, including 523 edges for vagino-cervical microbiome, 11,775 for fecal microbiome, 321 for metabolites, 280 for medical test, 70 for physical fitness test, 137 for facial skin feature, and 319 for lifestyle questionnaire. +,  $0.05 \leq Q < 0.1$ ; \*,  $Q < 0.05$ ; \*\*,  $Q < 0.01$ ; \*\*\*,  $Q < 0.001$ ; \*\*\*\*,  $Q < 0.0001$ ; \*\*\*\*\*,  $Q < 0.00001$ .

*vaginalis*, *L. johnsonii*, and *Weissella* species, and lower levels of plasma androstenedione, testosterone, and serum LDL (Figure 6; Table S6). Women with a heavier flow showed relatively more abundant *Propionibacterium acnes* in the cervical sample, as well as higher plasma manganese (Mn) and cobalt (Co) levels, but reported less problems of constipation and alopecia (Figure 6; Table S6). For the gut microbiome, we have recently showed association between gut microbial functional potential for metabolism of secondary bile acids and frequency of defecation [31]. Consistent with the association with a heavier flow, *P. acnes* has been found to be higher in the endometrium in the secretory phase than the proliferative phase in our previous study on surgical samples [8], and has previously been identified in the placenta and cultured from the follicular fluid [32–35]. Many women experience dysmenorrhea during menses (No:  $n = 84$ , Slight:  $n = 360$ , Serious:  $n = 68$ ) (Table S1). Individuals with dysmenorrhea were enriched for Pseudomonadales, *Acinetobacter*, and Moraxellaceae, while lower in the plasma level of histidine (Figure 5I), consistent with these bacteria encoding histidine decarboxylases to convert histidine into histamine [36]. Individuals with dysmenorrhea preferred spicy food instead of plain food (Figure 5I; Table S6). Thus, every aspect of the menstrual cycle could be seen in the vagino-cervical microbiome, together with measurements of hormones, amino acids, and rare elements in circulation.

#### Association between multi-omes

Integrated association network using a wisdom of crowds approach [37] also revealed interesting patterns (Spearman's correlation, random forest, and linear regression, with arcsin square root-transform for the microbiome profiles) (Figure 6; Table S6). Psychological questionnaire data were available from the first cohort (Figure 1). The relative abundance of *G. vaginalis* was negatively associated with sleep and diet assessment, while positively associated with plasma levels of mercury (Hg) and vitamin E (Figure 6; Table S6). Note that the cerebrospinal fluid flows from brain arteries to veins during deep sleep [38]. TTV (*Anellovirus* genus), better known for its immune surveillance function in the serum [39], was found here in cervical samples to be positively associated with oral contraceptives, negative emotions, and pores on the forehead, and negatively associated with WBC counts, corticosterone, and 11-deoxycortisol (Figure 6; Table S6). *L. iners* was negatively correlated with spontaneous abortion, days since last menstrual bleeding, condom usage, and the plasma concentration of vitamin A, and positively correlated with the plasma concentrations of hemoglobin and alanine (Figure 6; Table S6).

While lacking significant associations with the vagino-cervical microbial composition, the immune repertoire data, especially TRBV7.3 and TRBV7.4, showed multiple associations with functional pathways in the vagino-cervical microbiome, such as purine and pyrimidine metabolism, and synthesis of branched-chain amino acids, histidine, and arginine (Figure S8). Red blood cell (RBC) counts, plasma vitamin A levels, and plasma hydroxyl vitamin D levels were negatively associated with CDP-diacylglycerol biosynthesis pathways, consistent with the presence of diacylglycerol kinase in lactobacilli with anti-inflammatory functions [40,41]. Fecal *Coprococcus comes*, a bacterium previously reported to associate

with cytokine response to *C. albicans* [42], was seen here to be associated with isoleucine pathways in the vagino-cervical microbiome (Figure S8). Plasma levels of rare elements including arsenic (As) and mercury (Hg) also showed associations with functional pathways in the vagino-cervical microbiome. Both As and Hg were negatively associated with *de novo* synthesis, salvage, and degradation of purine. As negatively associated with the degradation of pyrimidine deoxyribonucleosides, Hg was negatively associated with biosynthesis of methionine and S-adenosyl-L-methionine (Figure S8). These results underscore the metabolic potential of the vagino-cervical microbiome, which should be sampled in addition to the fecal microbiome [31].

The second cohort had 263 salivary metagenomic shotgun data (Figure 1), which allowed us to explore the potential contribution from the oral microbiome. Oral *Eikenella corrodens* and unnamed *Comamonas* species were positively correlated with *G. vaginalis* in vagina (Table S7). *Scardovia wiggsiae*, a bacterium previously reported to be associated with early childhood caries, showed a positive correlation with vaginal *Staphylococcus* species (Figure S9; Table S7). Oral *Treponema lecithinolyticum* showed a negative correlation with *Dialister microaerophilus* (Figure S9; Table S7). Oral *Prevotella baroniae* was negatively correlated with *L. crispatus* in vagina (Figure S9; Table S7). The *NimI* gene, which exhibits high-level resistance to metronidazole, was previously reported to be intrinsic to *P. baroniae* [43].

We next analyzed the integrated association network between multi-omes in the second cohort and confirmed associations that were observed in both cohorts (Figure S9; Table S8). Patterns consistent with the initial cohort could be identified, such as negative correlations between pregnancy history, current breastfeeding, and lactobacilli (*L. crispatus*, *L. jensenii*, *L. iners*, and *L. vaginalis*), higher dehydroepiandrosterone, androstenedione, and testosterone with *L. crispatus* (Figure S9; Table S8). *Prevotella buccalis* and *Prevotella timonensis* were negatively correlated with the number of vaginal deliveries and the physical fitness test score, but positively correlated with the number of cesarean sections. *Finegoldia magna* was negatively correlated with vitamin B1 and Hg; *U. parvum* was positively correlated with arginine, serine, and threonine, while negatively correlated with 25-hydroxy vitamin D, 17 $\alpha$ -hydroxyprogesterone, and carnosine. *S. agalactiae* was negatively correlated with vitamin E, creatinine, and plasma concentration of hemoglobin. *P. bivia* was negatively correlated with 17 $\alpha$ -hydroxyprogesterone, 11-deoxycorticosterone, progesterone, and creatinine. Individuals who enriched for *L. iners* ranked better in wrinkles and red area on the forehead. Both *A. vaginae* and *G. vaginalis* were positively correlated with the numbers of RBCs (Figure S9; Table S8).

A total of 779 volunteers (51.99% from the initial cohort, 48.01% from the second cohort) had corresponding fecal metagenomic shotgun data. Among vaginal-cervical microbes, *L. crispatus* showed the strongest positive correlation with fecal microbes (pair-wise associations for all), which was also *L. crispatus* (the initial cohort,  $\rho = 0.386$ ,  $P = 9.47E-15$ ; the second cohort  $\rho = 0.326$ ,  $P = 2.24E-11$ ) (Figure 6, Figure S9; Table S8). Fecal *L. crispatus* was also positively correlated with *L. vaginalis* in the cervical microbiome (Table S8). The BV species *P. bivia* was negatively associated with fecal *Butyrivibrio*

Table 1 Summary the associated features for the major species in the vagino-cervical microbiome

Bacterium	Associated feature	Associated feature type	Potential relevance
<i>Lactobacillus crispatus</i>	Higher in the secretory phase vs. the proliferative phase Single, or without previous pregnancy, or without previous delivery Not currently breastfeeding Vaginal sexual debut at younger ages Less UV spots on forehead and cheeks Higher dehydroepiandrosterone, androstenedione, aldosterone, and testosterone Higher β-alanine Higher fecal <i>Lactobacillus crispatus</i> Without previous pregnancy, or lack of spontaneous abortion Higher dehydroepiandrosterone and corticosterone Not currently breastfeeding Higher in the proliferative phase vs. the secretory phase Oral contraceptive instead of condom Less wrinkles on forehead and spots on cheeks Higher alanine Higher plasma concentrations of hemoglobin Lower 25-hydroxy vitamin D3 Condom usage Without previous pregnancy Not currently breastfeeding Higher fecal <i>Haemophilus</i> species in individuals of the <i>Lactobacillus jensenii</i> type than in individuals of the <i>Lactobacillus crispatus</i> and <i>Atopobium vaginae</i> types Less spots on the cheek than women with <i>Lactobacillus crispatus</i> type Lower plasma androstenedione and testosterone Lower serum LDL Without previous pregnancy Not currently breastfeeding Young age Irregular menstruation RDW-SD Higher toughness assessment, psychological elasticity Vaginal douching Lower grip strength score Higher fecal <i>Prevotella hivia</i> Lower creatinine Lower 17 $\alpha$ -hydroxyprogesterone, 11-deoxycorticosterone, and progesterone Lower creatinine Number of caesarean sections instead of vaginal deliveries Lower comprehensive score of physical fitness test, lower vertical jump score Number of caesarean sections instead of vaginal deliveries Lower comprehensive score of physical fitness test, lower vertical jump score Shorter menstrual period Without marriage Dysmenorrhea Not oral probiotics within 1 month Longer menstrual period Lower vitamin E Lower creatinine and plasma concentrations of hemoglobin More serious problem of porphyrins and texture on the forehead Higher arginine, serine, and threonine; lower carnosine Lower 17 $\alpha$ -hydroxy progesterone Oral contraceptives Heavier flow in menstrual period Self-reported bacterial vaginosis Heavier flow in menstrual period Red area on the forehead Lower creatinine Lower grip strength score	Menstruation Marriage/pregnancy/delivery Breastfeeding Sexual debut Facial skin imaging Hormones Amino acids Fecal microbiome Pregnancy Hormones Breastfeeding Menstruation Contraception Facial skin imaging Amino acids Routine blood test Vitamin Contraception Pregnancy Breastfeeding Fecal microbiome Facial skin imaging Hormones Routine blood test Pregnancy Breastfeeding Age Menstruation Routine blood test Mental health conditions Vaginal douching Fitness training Fecal microbiome Routine blood test Hormones Routine blood test Delivery Fitness training Delivery Fitness training Menstruation Marriage Menstruation Probiotics Menstruation Vitamin Routine blood test Facial skin imaging Amino acids Vitamin Hormones Contraception Menstruation BV Menstruation Facial skin imaging Routine blood test Fitness training	Greater risk of BV resurgence after menses No such information available from iHMP when comparing ethnic groups [25] Better recovery of the microbiome after delivery Causal mechanism needed, e.g., testosterone associated with <i>L. crispatus</i> has been associated with life-time number of sexual partners  Detected in placenta, although controversial [10,32,71] Better recovery of the microbiome after delivery  HIV susceptibility in African countries [72] Early, more effective treatment for a young look  Contraception options Better recovery of the microbiome after delivery  Better recovery of the microbiome after delivery  BV or other infection Enhance physical fitness to improve the microbiome  Benefits of vaginal deliveries Enhance physical fitness to improve the microbiome Benefits of vaginal deliveries Enhance physical fitness to improve the microbiome  Effective treatment for dysmenorrhea Benefits of oral probiotics  Early, more effective treatment for a young look  Risk of oral contraceptives Bacterium from endometrium and fallopian tube [8] BV risk Phage or other treatment if really necessary Early, more effective treatment for a young look Muscle training
<i>Lactobacillus iners</i>			
<i>Lactobacillus gasseri</i> <i>Lactobacillus jensenii</i>			
<i>Lactobacillus vaginalis</i>			
<i>Gardnerella vaginalis</i>			
<i>Prevotella hivia</i>			
<i>Prevotella timonensis</i>			
<i>Prevotella buccalis</i>			
<i>Comamonas testosteroni</i>			
<i>Streptococcus agalactiae</i>			
<i>Ureaplasma parvum</i>			
<i>Propionibacterium acnes</i>			
<i>Dialister microaerophilus</i>			

Note: BV, bacterial vaginosis; iHMP, integrative Human Microbiome Project; UV, ultraviolet; HIV, human immunodeficiency virus; LDL, low density lipoprotein; RDW-SD, red cell distribution width-standard deviation.

species and *Clostridia* species, while positively associated with fecal *P. bivia* (the initial cohort,  $\rho = 0.115$ ,  $P = 0.021$ ; the second cohort,  $\rho = 0.097$ ,  $P = 0.059$ ) (Figure S9; Table S8). We were able to assemble *P. bivia* from one fecal sample, and sequences from its corresponding vagino-cervical sample also supported the genome (Figure S10A). On the other hand, *L. crispatus* could be assembled from vagino-cervical data, while its coverage in fecal metagenome was typically low (Figure S10B). Fecal *Porphyromonas somerae* was positively correlated with *P. timonensis* in the vagino-cervical microbiome in both cohorts (Table S8). More cases would need to be followed to conclude whether cervical *L. crispatus* translocates to the gut, and whether fecal *P. bivia* translocates to the reproductive tract, or *vice versa*. Together, we summarized the associated features for the major species in the vagino-cervical microbiome (Table 1), which could be further investigated in future cohorts.

## Discussion

As the largest metagenomic shotgun study for the vagino-cervical microbiome, our data revealed less known subtypes for the vagino-cervical microbiota, as well as detecting fungal and viral sequences. Our multi-omic data could help target efforts aimed at promoting a healthy reproductive tract microbiota and offering better advices for mothers from pregnancy to recovery, as well as preventing infections from viruses such as HIV and HPV. For example, whether vitamin D supplementation should be considered in African countries to rise *L. crispatus*, as investigated recently in a cohort of pregnant women [44]. Gut probiotics such as *Lactobacillus casei* and *Bifidobacterium longum* have been reported to increase vitamin D in ovariectomy-induced mice model of osteoporosis [45], and our results imply that similar bacteria in the vagina might also be related to vitamin D metabolism.

During the course of the initial review process, publications from integrative Human Microbiome Project (iHMP) provided longitudinal data during pregnancy and showed relatively more *L. iners* compared to non-pregnant individuals even in women of African American history [16,23,25,46]. In our data, the *L. iners* vs. *L. crispatus* shift was apparent in women with past pregnancy, and interestingly, *L. iners* showed a negative association with the number of spontaneous abortions (Figure 6; Table 1). *L. iners* and other lactobacilli have been detected in the placenta [10,11,32,47,48], amniotic fluid, nasal, and pharyngeal sites [49,50]. Aldosterone, a major mineralocorticoid for which we observed an association with potentially beneficial bacteria in the gut microbiome in another study [31], was positively associated with *L. crispatus* in the vagino-cervical microbiome; on the other hand, a precursor for aldosterone, corticosterone, was positively associated with *L. iners* (Figure 6; Table 1). How vitamin D and hormone metabolism might impact the vagino-cervical microbiome would require further studies. How the uterus and the microbiome recover during breastfeeding is also of interest for both the mother and future children [16,51]. *S. anginosus* has been shown to be more abundant in the gut microbiome of individuals with atherosclerotic cardiovascular diseases [52], and here we tentatively observed young women who harbored > 50% *S. anginosus* in cervical samples showing higher plasma creatinine and lower 17 $\alpha$ -hydroxyprogesterone (Figure 2), which

might relate to preterm birth [10]. The vagino-cervical microbiome of postmenopausal women revealed a myriad of viral and bacterial species (Figure S5), while fungal growth might be unfavorable both due to high pH and lack of glycans. While a focus in the field of the female vaginal microbiota has been infection and preterm birth, our data highlight major aspects of importance for female health that are worth further investigations for women in the modern world.

Samples from multiple body sites (peritoneal fluid from the pouch of Douglas, fallopian tubes, endometrium, cervical mucus, and two vaginal sites) have been analyzed from volunteers with benign conditions such as hysteromyoma, adenomyosis, and endometriosis [8,9]. Here we were only able to sample the vagino-cervical microbiome in this healthy cohort. Interestingly, our cohort generally lacked bacteria known for BV (e.g., studies from Serrano et al. [25] and Fettweis et al. [46]) and also contained less *Prevotella* compared to some of the surgically sampled individuals. In the postmenopausal samples, some of the bacteria previously reported in the upper reproductive tract [8,9,53] and involved in degradation of hormones [54], e.g., *Pseudomonas* spp., could be seen in the vagino-cervical microbiome, while species of disease relevance such as *A. vaginae* and *Porphyromonas* in endometrial cancer would require more evidence [12].

Sampling of the cervix with a cytobrush by experienced doctors allowed the analyses of the microbiome that is generally lactobacilli-dominated like the vaginal sites of the lower reproductive tract [8,9]. Cervix samples could also reflect the individual-specific continuum of the microbiome from the upper reproductive tract including the peritoneal cavity [8,9]. We observed here that the association with the fecal microbiome was limited, and the notion of reservoirs in the intestine or other sites for vagino-cervical bacteria such as *L. crispatus* and *P. bivia* would need further investigation, especially in light of individual differences in the number of CD4<sup>+</sup> T cells in mucosal sites [17,55]. We observed other interesting associations between different species potentially involved in immune modulation (Figure 6). The plasma metabolites and T cell receptor types associated with the vagino-cervical microbiome were distinct from those associated with the fecal microbiome [31]. Vaginal *Prevotella* could induce more CD4<sup>+</sup> T cells [17], but the *Prevotella copri* was not the dominant gut species of *Prevotella* species which may compete with *Bacteroides* spp. [52]. The vagino-cervical microbiome also better predicted facial skin features compared to the fecal microbiome [31], perhaps due to a clearer pattern of hormone and immune signatures. The associations with physical fitness tests and self-reported physical activities were, however, less prominent in the vagino-cervical microbiome compared to the fecal microbiome [31], as the changes due to pregnancy, delivery, and breastfeeding may not be easily modifiable with physical activity. Interesting associations were identified between the vagino-cervical microbiome and physical fitness test results, e.g., *P. bivia* was negatively associated with hand grip strength, and *P. buccalis*, *Prevotella disiens*, and *Peptoniphilus harei* were negatively associated with vertical jump score. As a densely populated microbiota other than the distal gut, the vagino-cervical microbiota has the potential to reflect or even influence physiology elsewhere in the human body.

## Materials and methods

### Initial study cohort

As the first time point for the vagino-cervical microbiome of the 4D-SZ cohort, 516 Chinese volunteers joined from May to July during an annual physical examination in 2017. Baseline characteristics of the cohort are shown in Table S1.

### The second cohort

An independent cohort of 632 Chinese volunteers were recruited from May to July during an annual physical examination in 2018. 2018 data for 4D-SZ volunteers who were already included in the initial cohort were excluded and will be published in a future study. The collection procedures of samples and multi-omic data were similar to that in the initial cohort. In addition, salivary samples were collected only in this cohort. Baseline characteristics of the cohort are shown in Table S1.

### Demographic data collection

During physical examination, the volunteers received three kinds of online questionnaires. 1) The female life history questionnaire contained pregnancy and delivery histories, menstrual phases, sexual activity, and contraceptive methods. 2) The lifestyle questionnaire contained age, disease history, and eating and exercise habits. 3) The psychological questionnaire contained the evaluation of irritability, dizziness, frustration, fear, appetite, self-confidence, and resilience (Table S1).

### Sample collection

Cervical samples were collected and smeared in the Flinders Technology Associates (FTA) cards by doctor during gynecological examination. Fecal samples and salivary samples were self-collected by volunteers. Cervical samples, fecal samples, and salivary samples were stored at  $-80^{\circ}\text{C}$  for metagenomic shotgun sequencing. The overnight fasting blood samples were drawn from a cubital vein of volunteers by doctor.

### DNA extraction and metagenomic shotgun sequencing

DNA extraction of cervical samples and fecal samples was performed as described [8,56]. Metagenomic shotgun sequencing was performed on the BGISEQ-500 platform, which is highly comparable to Illumina HiSeq platforms in metagenomic and other sequencing applications [9,57–59]. The 50 bp of single-end reads for cervical samples collected in the initial study cohort, and on an average of 208.76 million raw reads were sequenced for each sample (Table S2); the 100 bp of single-end reads for fecal samples collected in the initial study cohort, and on an average of 85.63 million raw reads were sequenced for each sample (Table S2); the 100 bp of paired-end reads for cervical samples, fecal samples, and salivary samples collected in the second cohort, and on an average of 158.91 million raw reads for cervical samples, 151.69 million raw reads for salivary samples, and 148.26 million raw reads for fecal samples were sequenced for each sample, respectively (Table S2). Quality

control and alignment to GRCh38 were performed as previously described [9,57].

### Ultra high pressure liquid chromatography-mass spectrometry quantification of amino acids

40  $\mu\text{l}$  of plasma was deproteinized with 20  $\mu\text{l}$  of 10% (w/v) sulfosalicylic acid (Sigma) containing internal standards, and then 120  $\mu\text{l}$  of 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 MS (AB Sciex, Los Angeles, CA) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.8  $\mu\text{m}$ , 2.1 mm  $\times$  100 mm) was used for amino compound separation with a flow rate at 0.5 ml/min and column temperature of 55  $^{\circ}\text{C}$ . The mobile phases were phase A [water containing 0.05% heptafluorobutyric acid and 0.1% formic acid (v/v)] and phase B [acetonitrile containing 0.05% heptafluorobutyric acid 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 rate was 35 l/min; Collision Gas (CAD) was medium; Ion Source Gas 1 (GS 1) flow rate was 60 l/min; Ion Source Gas 2 (GS 2) flow rate was 60 l/min; IonSpray Voltage (IS) was 5500 V and temperature was 600  $^{\circ}\text{C}$ . All amino standard reagents were purchased from Sigma-Aldrich (St. Louis, MO) and Toronto research chemical (TRC).

### UHPLC-MS quantification of hormones

250  $\mu\text{l}$  of plasma was diluted with 205  $\mu\text{l}$  of sterile water. For solid-phase extraction (SPE) experiments, hydrophilic lipophilic balance (HLB, Waters) was supplemented with 1.0 ml for each of dichloromethane, acetonitrile, and methanol, and was equilibrated with 1.0 ml of water. The pretreated plasma sample 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 transferred to an autosampler vial prior to UHPLC-MS analysis. The analysis was performed by UHPLC coupled to an AB Sciex Qtrap 5500 MS (AB Sciex) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Phenomene Kinetex C18 column (2.6  $\mu\text{m}$ , 2.1 mm  $\times$  50 mm) was used for steroid hormone separation with a flow rate at 0.8 ml/min and column temperature of 55  $^{\circ}\text{C}$ . The mobile phases were phase A (water containing 1 mM ammonium acetate) and phase B (methanol containing 1 mM ammonium acetate). The gradient elution was 25% B kept for 0.9 min, then changed linearly to 40% B during 0.9 min, continued up to 70% B in 2 min, increased to 95% B in 0.1 min, and maintained for 1.6 min. MRM was used to monitor all steroid hormone compounds. The mass parameters were as follows: curtain gas flow rate was 35 l/min; CAD was medium; GS 1 flow rate was

60 l/min; GS 2 flow rate was 60 l/min; Nebulizer Current (NC) was 5 and temperature was 500 °C. All steroid hormone profiling compound standards were purchased from Sigma-Aldrich, TRC, Cerilliant, and Dr. Ehrenstorfer.

#### **Inductively coupled plasma-mass spectrometry quantification of trace elements**

200 µl of whole blood was transferred into a 15-ml polyethylene tube and diluted 1:25 with a diluent solution consisting of 0.1% (v/v) Triton X-100, 0.1% (v/v) HNO<sub>3</sub>, 2 mg/l AU, and internal standards (20 µg/l). The mixture was sonicated for 10 min before inductively coupled plasma-mass spectrometry (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-mm diameter injector with a Shield Torch system, and a Scott double-pass spray chamber and nickel cones (Agilent Technologies). A glass concentric MicroMist nebuliser (Agilent Technologies) was used for the analysis of diluted samples.

#### **Ultra pressure liquid chromatography-mass spectrometry quantification of water-soluble vitamins**

200 µl of plasma was deproteinized with 600 µl of methanol (Merck), water, and 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 of supernatant was dried under a nitrogen flow. 60 µl of water was added to the residues, and then vortexed. The mixture was centrifuged and the supernatant was used for analysis. The analysis was performed by ultra pressure liquid chromatography (UPLC) coupled to a Waters Xevo TQ-S Triple Quad MS (Waters) with the ESI source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.7 µm, 2.1 mm × 50 mm) was used for water-soluble vitamin separation with a flow rate at 0.45 ml/min and column temperature of 45 °C. The mobile phases were phase A (0.1 % formic acid in water) and phase B (0.1% formic acid in methanol). The following elution gradient was used: 0–1 min, 99% A; 1–1.5 min, 99%–97% A; 1.5–2 min, 97%–70% A; 2–3.5 min, 70%–30% A; 3.5–4 min, 30%–10% A; 4–4.8 min, 10% A; 4.9–6 min, 99% A. MRM was used to monitor all water-soluble vitamins. The mass parameters were as follows: the capillary voltages of 3000 V and source temperature of 150 °C were adopted. The desolvation temperature was 500 °C. The CAD flow rate was set at 0.10 ml/min. The cone gas and desolvation gas flow rates were 150 l/h and 1000 l/h, respectively. All water-soluble vitamin standards were purchased from Sigma-Aldrich.

#### **UPLC-MS quantification of fat-soluble vitamins**

250 µl of plasma was deproteinized with 1000 µl of 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, and α-Tocopherol-d6 (Toronto Research Chemicals). 900 µl of supernatant was dried under a nitrogen flow. 80 µl of 80% acetonitrile was added to the residues, and then vortexed. The mixture was centrifuged, and the supernatant was used for analysis. The analysis was performed by UPLC coupled to an AB Sciex Qtrap 4500 MS (AB Sciex) with the APCI source in positive ion mode. A Waters ACQUITY UPLC BEH C18 column (1.7 µm, 2.1 mm × 50 mm) was used for fat-soluble vitamin separation with a flow rate at 0.50 ml/min and column temperature of 45 °C. The mobile phases were phase A (0.1 % formic acid in water) and phase B (0.1% formic acid in acetonitrile). The following elution gradient was used: 0–0.5 min, 60% A; 0.5–1.5 min, 60%–20% A; 1.5–2.5 min, 20%–0% A; 2.5–4.5 min, 0% A; 4.5–4.6 min, 0%–60% A; 4.6–5.0 min, 60% A. MRM was used to monitor all fat-soluble vitamins. The mass parameters were as follows: curtain gas flow rate was 30 l/min; CAD was medium; GS 1 flow rate was 40 l/min; GS 2 flow rate was 50 l/min; NC was 5 and the temperature was 400 °C. All fat-soluble vitamin standards were purchased from Sigma-Aldrich and TRC.

#### **Sequencing of the T-cell receptor β complementarity-determining region 3 immune repertoire**

10 ml of whole blood was centrifuged at 3000 r/min for 10 min, and then 165 µl of buffy coat was obtained to extract DNA using MagPure Buffy Coat DNA Midi KF Kit (Magen, China). The DNA was sequenced on the BGISEQ-500 platform using 200 bp single-end reads. Data processing was performed using Immune IMonitor [60].

#### **Medical parameters**

All the volunteers were recruited during the physical examination. The medical test included blood tests, urinalysis, and routine examination of cervical secretion. InBody was measured by InBody Analyzer (InBody Co., Ltd). All the medical parameters were measured by the physical examination center and shown in Table S1.

#### **Facial skin features**

The volunteers were required to clean their face and use no makeup after they got up in the morning. The volunteer's frontal face was photographed by VISIA-CRTM imaging system (Canfield Scientific, Fairfield, NJ) 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 (Table S1). The percentile for each index was calculated based on the index value ranked in the age-matched database (Table S1). The higher the percentile of an index, the better the facial skin appears.

#### **Physical fitness test**

Eight kinds of tests were performed to evaluate the volunteers' physical fitness condition (Table S1). Vital capacity was

measured by HK6800-FH (Hengkangjiaye, China). Single-legged stance with eyes closed was measured by HK6800-ZL. Choice reaction time was measured by HK6800-FY. Grip strength was measured by HK6800-WL. Sit-and-reach score was measured by HK6800-TQ. Sit-ups were 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.

### Quality control, taxonomic annotation, and abundance calculation

The sequencing reads were quality-controlled by the overall accuracy (OA) control strategy as described previously [57]. Reads with length no short than 30 bp, seed OA higher than 0.9, and fragment OA higher than 0.8 were retained.

Host sequence contamination reads in fecal and salivary samples were removed using soap2.22 and hg19 index with parameters “-M 4 -l 30 -r 1 -m 200 -x 600 -v 8 -c 0.9”. As there were over 90% of human sequences in cervical samples, the stringent condition for removal of host sequences was used [9]. In short, after using the same SOAP2 process, the retained reads were mapped to hg19, hg38, and YH indexes by Decon-Seq [61] and SNAP1.0 [62], respectively. Unaligned reads were retained for analysis.

Taxonomic assignment of the high-quality cervical metagenomic shotgun data and salivary metagenomic shotgun data was performed using MetaPhlan2 [63].

Taxonomic assignment of the high-quality fecal metagenomic shotgun data was performed using the reference gene catalog comprising 9,879,896 genes [56]. Taxonomies of the fecal metagenomic linkage groups / metagenomic species were then determined from their constituent genes, as previously described [4,64,65].

The vaginal samples were hierarchically clustered using R base hcluster function with centroid linkage based on Euclidean distance in Figure 2A.

### Random forest on the influence of female life history factors

The factors in the female life history questionnaire were fitted against the relative abundances of taxonomic profiles (found in at least 10% of the samples) of the cervical samples using default parameters in the RFCV regression function from randomForest package in R. Female life history factors, except age (a continuous variable), are either dummy variables such as pregnancy history (yes, no) or frequency variables such as number of caesarean sections (0, 1, 2) (Table S1). In addition to comparing the predictive power across factors, we used regression model instead of classification model here. Spearman’s correlation between the measured value and the 5-fold cross-validation predicted value was calculated as a model performance metric. Then the key predictable factors were ranked. *P* value was obtained using permutation test (999 times).

### The global effect size between vagino-cervical microbiome and omic data

To evaluate the combined effect size of vagino-cervical microbiome on omic data, we used forward stepwise redundancy

analysis of omic data lists on the relative abundances of taxonomic profiles in forward.sel function in the packfor package in R. This analysis provided a global vs. global association between any two omic datasets that maximizes the associations using the strongest predictive power of non-redundant predictors.

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

The factors in each type of omes were fitted against the relative abundances of taxonomic profiles (found in at least 10% of the samples) of the cervical samples using default parameters in the RFCV regression function from randomForest package in R. Omic data are a mix of dummy variables and continuous variables (Table S1). In addition to comparing the predictive power across factors, we used regression model instead of classification model here. Spearman’s correlation between the measured value and the 5-fold cross-validation predicted value was calculated as a model performance metric. Then the top 8 predictable factors in each type were ranked. *P* value was obtained using permutation test (999 times).

### Transformation of metagenomic shotgun profiles for composition data analysis

We normalized the microbiome data with arcsine square root transformation to make it less skewed for the downstream analysis (implemented in MaAsLin software [66]).

### Wisdom of crowds for robust network construction between vagino-cervical microbial species and multi-omic data

A new method for multi-omics analyses [37] was used to integrate the coefficient of linear regression, variance importance from randomForest, and Spearman’s correlation to construct omic flux networks, and then visualized the networks in Cytoscape. The details are as indicated below.

Step 1: Data processing. All categorical variables in multi-omic data were converted into continuous variables, and nominal variables were converted into dummy variables. Missing values were filled with median, and samples containing more than 70% missing variables were removed. The microbial species less than 10% in all the samples were also removed, as well as the 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 [67]. RFCV regression function from randomForest package in R with default parameters 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, and then bootstrapping glmnet with 0.632 re-sampling was performed 100 times to get 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 (Figure 5). Cytoscape was also used to visualize the omic network (Figure 6).

The second cohort was analyzed using the same statistical method. Combining  $P$  value is computed using Edgington method from metap package in R. BH method was used to adjust the multiple test  $P$  value ( $Q$  value). The correlation is identified based on  $Q < 0.1$  when a similar microbial distribution pattern shown in the initial study cohort and the second cohort.

### Association between microbiome pathways and multi-omes

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

### Circular genome map of *P. bivia* and *L. crispatus*

The high-quality reads of fecal and vaginal samples from two separate individuals with high amounts of *L. crispatus* or *P. bivia* were singled out to assemble using metaSPAdes (with the parameter “spades.py --meta -t 8 -m 50”). Then Blastn (with the parameter “blastn -word\_size 16 -outfmt 6 -evalue 1e-10 -max\_target\_seqs 5000 -num\_threads 8”) was used to extract the contigs of corresponding species. The read depth of the contigs was aligned to the RefSeq Database (2019-06-06) with the parameter “bwa mem -t 8”. The comparison of the vaginal and fecal assembled genomes from the same individual was visualized by BRIG (<https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-402>).

### Ethical statement

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

### Data availability

Metagenomic shotgun data for all samples and other relevant data reported in this study have been deposited in both CNGB Sequence Archive [68] of China National GeneBank DataBase (CNGBdb) [69] (CNSA: CNP0000287), and the Genome Warehouse [70] at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformatics (BioProject: PRJCA003712) which are publicly accessible at <https://ngdc.cncb.ac.cn/gsa/>.

### CRedit author statement

**Zhuye Jie:** Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Chen Chen:** Conceptualization, Investigation, Visualization, Writing - original draft, Project administration. **Lilan Hao:** Investigation, Methodol-

ogy, Formal analysis, Visualization, Writing - original draft. **Fei Li:** Formal analysis, Visualization. **Liju Song:** Investigation. **Xiaowei Zhang:** Writing - original draft. **Jie Zhu:** Formal analysis. **Liu Tian:** Formal analysis. **Xin Tong:** Investigation. **Kaiye Cai:** Investigation. **Zhe Zhang:** Writing - original draft. **Yanmei Ju:** Investigation. **Xinlei Yu:** Investigation. **Ying Li:** Investigation. **Hongcheng Zhou:** Investigation. **Haorong Lu:** Investigation. **Xuemei Qiu:** Investigation. **Qiang Li:** Investigation. **Yunli Liao:** Investigation. **Dongsheng Zhou:** Investigation. **Heng Lian:** Investigation. **Yong Zuo:** Investigation. **Xiaomin Chen:** Investigation. **Weiqiao Rao:** Investigation. **Yan Ren:** Investigation. **Yuan Wang:** Investigation. **Jin Zi:** Investigation. **Rong Wang:** Investigation. **Na Liu:** Investigation. **Jinghua Wu:** Investigation. **Wei Zhang:** Investigation. **Xiao Liu:** Investigation. **Yang Zong:** Investigation. **Weibin Liu:** Investigation. **Liang Xiao:** Supervision. **Yong Hou:** Supervision. **Xun Xu:** Supervision. **Huanming Yang:** Supervision. **Jian Wang:** Supervision. **Karsten Kristiansen:** Writing - review & editing. **Huijue Jia:** Conceptualization, Writing - review & editing, Supervision, Project administration. All authors have read and approved the final manuscript.

### Competing interests

The authors have declared no competing interests.

### Acknowledgments

We are very grateful to colleagues at BGI-Shenzhen, China for sample collection and discussions, and China National GeneBank (CNGB) Shenzhen for DNA extraction, library construction, and sequencing. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2021.01.005>.

### ORCID

ORCID 0000-0003-0833-6134 (Zhuye Jie)  
 ORCID 0000-0001-6788-2894 (Chen Chen)  
 ORCID 0000-0003-0198-063X (Lilan Hao)  
 ORCID 0000-0002-8142-1100 (Fei Li)  
 ORCID 0000-0001-9935-181X (Liju Song)  
 ORCID 0000-0002-2446-7487 (Xiaowei Zhang)  
 ORCID 0000-0002-5789-6296 (Jie Zhu)  
 ORCID 0000-0002-4421-5105 (Liu Tian)  
 ORCID 0000-0002-4849-6522 (Xin Tong)  
 ORCID 0000-0003-0593-1878 (Kaiye Cai)  
 ORCID 0000-0003-3983-1226 (Zhe Zhang)  
 ORCID 0000-0002-1600-3164 (Yanmei Ju)  
 ORCID 0000-0003-2664-9388 (Xinlei Yu)  
 ORCID 0000-0001-5131-2803 (Ying Li)  
 ORCID 0000-0003-4644-6940 (Hongcheng Zhou)  
 ORCID 0000-0003-0132-8166 (Haorong Lu)

ORCID 0000-0002-4406-6728 (Xuemei Qiu)  
 ORCID 0000-0002-6099-8883 (Qiang Li)  
 ORCID 0000-0001-6472-3847 (Yunli Liao)  
 ORCID 0000-0003-0594-0117 (Dongsheng Zhou)  
 ORCID 0000-0001-8251-7473 (Heng Lian)  
 ORCID 0000-0002-8463-4267 (Yong Zuo)  
 ORCID 0000-0003-0046-895X (Xiaomin Chen)  
 ORCID 0000-0002-2738-672X (Weiqiao Rao)  
 ORCID 0000-0002-4007-8625 (Yan Ren)  
 ORCID 0000-0001-8071-2451 (Yuan Wang)  
 ORCID 0000-0001-9224-0199 (Jin Zi)  
 ORCID 0000-0003-4565-8925 (Rong Wang)  
 ORCID 0000-0002-9088-6794 (Na Liu)  
 ORCID 0000-0002-0883-753X (Jinghua Wu)  
 ORCID 0000-0002-6968-6974 (Wei Zhang)  
 ORCID 0000-0002-8073-0534 (Xiao Liu)  
 ORCID 0000-0002-0690-5439 (Yang Zong)  
 ORCID 0000-0002-9349-039X (Weibin Liu)  
 ORCID 0000-0003-0836-4397 (Liang Xiao)  
 ORCID 0000-0002-0420-0726 (Yong Hou)  
 ORCID 0000-0002-5338-5173 (Xun Xu)  
 ORCID 0000-0002-0858-3410 (Huanming Yang)  
 ORCID 0000-0003-4949-4753 (Jian Wang)  
 ORCID 0000-0002-6024-0917 (Karsten Kristiansen)  
 ORCID 0000-0002-3592-126X (Huijue Jia)

## References

- [1] Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. *Cell* 2016;164:337–40.
- [2] Sommer F, Bäckhed F. The gut microbiota — masters of host development and physiology. *Nat Rev Microbiol* 2013;11:227–38.
- [3] O'Toole PW, Marchesi JR, Hill C. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nat Microbiol* 2017;2:17057.
- [4] Wang J, Jia H. Metagenome-wide association studies: fine-mining the microbiome. *Nat Rev Microbiol* 2016;14:508–22.
- [5] Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.
- [6] Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 2017;550:61–6.
- [7] Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 2015;21:895–905.
- [8] Chen C, Song X, Wei W, Zhong H, Dai J, Lan Z, et al. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat Commun* 2017;8:875.
- [9] Li F, Chen C, Wei W, Wang Z, Dai J, Hao L, et al. The metagenome of the female upper reproductive tract. *Gigascience* 2018;7:giy107.
- [10] de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, et al. Human placenta has no microbiome but can contain potential pathogens. *Nature* 2019;572:329–34.
- [11] Seferovic MD, Pace RM, Carroll M, Belfort B, Major AM, Chu DM, et al. Visualization of microbes by 16S *in situ* hybridization in term and preterm placentas without intraamniotic infection. *Am J Obstet Gynecol* 2019;221:146.
- [12] Walther-António MRS, Chen J, Multinu F, Hokenstad A, Distad TJ, Cheek EH, et al. Potential contribution of the uterine microbiome in the development of endometrial cancer. *Genome Med* 2016;8:122.
- [13] Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, Mcculle SL, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 2011;108:4680–7.
- [14] Ma B, Forney LJ, Ravel J. Vaginal microbiome: rethinking health and disease. *Annu Rev Microbiol* 2012;66:371–89.
- [15] Petrova MI, van den Broek M, Balzarini J, Vanderleyden J, Lebeer S. Vaginal microbiota and its role in HIV transmission and infection. *FEMS Microbiol Rev* 2013;37:762–92.
- [16] DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* 2015;112:11060–5.
- [17] Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, Bowman BA, et al. *Lactobacillus*-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity* 2017;46:29–37.
- [18] Bradford LL, Ravel J. The vaginal mycobiome: a contemporary perspective on fungi in women's health and diseases. *Virulence* 2017;8:342–51.
- [19] Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012;486:215–21.
- [20] Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* 2005;353:1899–911.
- [21] Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. *Nat Rev Microbiol* 2018;16:143–55.
- [22] Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UME, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 2012;4:132ra52.
- [23] Petricevic L, Domig KJ, Nierscher FJ, Sandhofer MJ, Fidesser M, Krondorfer I, et al. Characterisation of the vaginal *Lactobacillus* microbiota associated with preterm delivery. *Sci Rep* 2014;4: 5136.
- [24] Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpidis NC, et al. Genome-based taxonomic classification of the phylum Actinobacteria. *Front Microbiol* 2018;9:2007.
- [25] Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, Edupuganti L, et al. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. *Nat Med* 2019;25: 1001–11.
- [26] Harriott MM, Lilly EA, Rodriguez TE, Fidel PL, Noverr MC. *Candida albicans* forms biofilms on the vaginal mucosa. *Microbiology* 2010;156:3635–44.
- [27] Horinouchi M, Hayashi T, Yamamoto T, Kudo T. A new bacterial steroid degradation gene cluster in *Comamonas testosteroni* TA441 which consists of aromatic-compound degradation genes for seco-steroids and 3-ketosteroid dehydrogenase genes. *Appl Environ Microbiol* 2003;69:4421–30.
- [28] Anderson BL, Mendez-Figueroa H, Dahlke JD, Raker C, Hillier SL, Cu-Uvin S. Pregnancy-induced changes in immune protection of the genital tract: defining normal. *Am J Obstet Gynecol* 2013;208:321.
- [29] Pattaroni C, Watzenboeck ML, Schneidegger S, Kieser S, Wong NC, Bernasconi E, et al. Early-life formation of the microbial and immunological environment of the human airways. *Cell Host Microbe* 2018;24:857–65.
- [30] MacNeil LT, Watson E, Arda HE, Zhu LJ, Walhout AJM. Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell* 2013;153:240–52.
- [31] Jie Z, Liang S, Ding Q, Li F, Tang S, Wang D, et al. A transomic cohort as a reference point for promoting a healthy human gut microbiome. *Medicine in Microecology* 2021;8:100039.
- [32] Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. *Sci Transl Med* 2014;6:237ra65.
- [33] Pelzer ES, Allan JA, Theodoropoulos C, Ross T, Beagley KW, Knox CL. Hormone-dependent bacterial growth, persistence and

- biofilm formation — a pilot study investigating human follicular fluid collected during IVF cycles. *PLoS One* 2012;7:e49965.
- [34] Pelzer ES, Harris JE, Allan JA, Waterhouse MA, Ross T, Beagley KW, et al. TUNEL analysis of DNA fragmentation in mouse unfertilized oocytes: the effect of microorganisms within human follicular fluid collected during IVF cycles. *J Reprod Immunol* 2013;99:69–79.
- [35] Pelzer ES, Allan JA, Waterhouse MA, Ross T, Beagley KW, Knox CL. Microorganisms within human follicular fluid: effects on IVF. *PLoS One* 2013;8:e59062.
- [36] Cundell DR, Devalia JL, Wilks M, Tabaqchali S, Davies RJ. Histidine decarboxylases from bacteria that colonise the human respiratory tract. *J Med Microbiol* 1991;35:363–6.
- [37] Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. *Nat Methods* 2012;9:796–804.
- [38] Nedergaard M, Goldman SA. Glymphatic failure as a final common pathway to dementia. *Science* 2020;370:50–6.
- [39] Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell* 2009;138:30–50.
- [40] Ganesh BP, Hall A, Ayyaswamy S, Nelson JW, Fultz R, Major A, 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* 2018;11:380–93.
- [41] Andrada E, Liébana R, Merida I. Diacylglycerol kinase  $\zeta$  limits cytokine-dependent expansion of CD8<sup>+</sup> T cells with broad antitumor capacity. *EBioMedicine* 2017;19:39–48.
- [42] Schirmer M, Smeeckens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, et al. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell* 2016;167:1125–36.
- [43] Alauzet C, Mory F, Teyssier C, Hallage H, Carlier JP, Grollier G, et al. Metronidazole resistance in *Prevotella* spp. and description of a new *nim* gene in *Prevotella baroniae*. *Antimicrob Agents Chemother* 2010;54:60–4.
- [44] Jefferson KK, Parikh HI, Garcia EM, Edwards DJ, Serrano MG, Hewison M, et al. Relationship between vitamin D status and the vaginal microbiome during pregnancy. *J Perinatol* 2019;39:824–36.
- [45] Montazeri-Najafabady N, Ghasemi Y, Dabbaghmanesh MH, Talezadeh P, Koohpeyma F, Gholami A. Supportive role of probiotic strains in protecting rats from ovariectomy-induced cortical bone loss. *Probiotics Antimicrob Proteins* 2019;11:1145–54.
- [46] Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The vaginal microbiome and preterm birth. *Nat Med* 2019;25:1012–21.
- [47] Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G. At the crossroads of vaginal health and disease, the genome sequence of *Lactobacillus iners* AB-1. *Proc Natl Acad Sci U S A* 2011;108:4688–95.
- [48] Lannon SMR, Adams Waldorf KM, Fiedler T, Kapur RP, Agnew K, Rajagopal L, et al. Parallel detection of *Lactobacillus* and bacterial vaginosis-associated bacterial DNA in the chorioamnion and vagina of pregnant women at term. *J Matern Neonatal Med* 2019;32:2702–10.
- [49] Wang J, Zheng J, Shi W, Du N, Xu X, Zhang Y, et al. Dysbiosis of maternal and neonatal microbiota associated with gestational diabetes mellitus. *Gut* 2018;67:1614–25.
- [50] de Boeck I, van den Broek MFL, Allonsius CN, Spacova I, Wittouck S, Martens K, et al. *Lactobacilli* have a niche in the human nose. *Cell Rep* 2020;31:107674.
- [51] Anton L, Sierra LJ, DeVine A, Barila G, Heiser L, Brown AG, et al. Common cervicovaginal microbial supernatants alter cervical epithelial function: mechanisms by which *Lactobacillus crispatus* contributes to cervical health. *Front Microbiol* 2018;9:2181.
- [52] Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun* 2017;8:845.
- [53] Lee S, La TM, Lee HJ, Choi IS, Song CS, Park SY, et al. Characterization of microbial communities in the chicken oviduct and the origin of chicken embryo gut microbiota. *Sci Rep* 2019;9:6838.
- [54] Wang P, Zheng D, Wang Y, Liang R. One 3-oxoacyl-(acyl-carrier-protein) reductase functions as 17 $\beta$ -hydroxysteroid dehydrogenase in the estrogen-degrading *Pseudomonas putida* SJTE-1. *Biochem Biophys Res Commun* 2018;505:910–6.
- [55] Marrazzo JM, Fiedler TL, Srinivasan S, Thomas KK, Liu C, Ko D, et al. Extravaginal reservoirs of vaginal bacteria as risk factors for incident bacterial vaginosis. *J Infect Dis* 2012;205:1580–8.
- [56] Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;32:834–41.
- [57] Fang C, Zhong H, Lin Y, Chen B, Han M, Ren H, et al. Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *Gigascience* 2018;7:1–8.
- [58] Han MM, Hao L, Lin Y, Li F, Wang J, Yang H, et al. A novel affordable reagent for room temperature storage and transport of fecal samples for metagenomic analyses. *Microbiome* 2018;6:43.
- [59] Pan H, Guo R, Zhu J, Wang Q, Ju Y, Xie Y, et al. A gene catalogue of the Sprague-Dawley rat gut metagenome. *Gigascience* 2018;7:giy055.
- [60] Zhang W, Du Y, Su Z, Wang C, Zeng X, Zhang R, et al. IMonitor: a robust pipeline for TCR and BCR repertoire analysis. *Genetics* 2015;201:459–72.
- [61] Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 2011;6:e17288.
- [62] Zaharia M, Bolosky WJ, Curtis K, Fox A, Patterson D, Shenker S, et al. Faster and more accurate sequence alignment with SNAP. *arXiv* 2011;1111.5572.
- [63] Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 2012;9:811–4.
- [64] Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 2014;32:822–8.
- [65] Qin J, Li Y, Cai Z, Li SS, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012;490:55–60.
- [66] Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79.
- [67] Louppe G, Wehenkel L, Sutura A, Geurts P. Understanding variable importances in forests of randomized trees. *Adv Neural Inf Process Syst* 2013;26:431–9.
- [68] Guo X, Chen F, Gao F, Li L, Liu K, You L, et al. CNSA: a data repository for archiving omics data. *Database* 2020;2020:baaa055.
- [69] Chen F, You L, Yang F, Wang L, Guo X, Gao F, et al. CNGBdb: China National GeneBank DataBase. *Hereditas* (Beijing) 2020;42:799–809. (in Chinese with an English abstract)
- [70] Chen M, Ma Y, Wu S, Zheng X, Kang H, Sang J, et al. Genome Warehouse: a public repository housing genome-scale data. *Genomics Proteomics Bioinformatics* 2021;19:584–9.
- [71] Parnell LA, Briggs CM, Cao B, Delannoy-Bruno M, Schrieffer AE, Mysorekar IU. Microbial communities in placentas from term normal pregnancy exhibit spatially variable profiles. *Sci Rep* 2017;7:11200.
- [72] Borgdorff H, Tsvitvadze E, Verhelst R, Marzorati M, Jurriaans S, Ndayisaba GF, et al. *Lactobacillus*-dominated cervicovaginal microbiota associated with reduced HIV/STI prevalence and genital HIV viral load in African women. *ISME J* 2014;8:1781–93.