Life history recorded in the vagino-cervical microbiome

Zhuye Jie[1,2,3,*], Chen Chen[1,3,4,*,†], Lilan Hao[1,3], Fei Li[1,3], Liju Song[1,3], Xiaowei Zhang[1,3], Liu Tian[1,3], Xin Tong[1,3], Kaiye Cai[1], Yanmei Ju[1,3], 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], Jinghua Wu[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†]

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

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

SUMMARY

The gut microbiome has been the center of attention for human commensal microbiome studies. For the vagina, another densely populated body site, 16S rRNA gene amplicon sequencing studies have identified community types and observed dynamic changes related to menstrual cycles and sexual behaviors in small cohorts. Here we perform metagenomic shotgun sequencing on cervical samples from 1148 volunteers (a cohort more than 20-fold of the Human Microbiome Project (HMP)). Factors such as pregnancy and delivery histories and contraception methods were all more important than menstrual cycle, and contributing to non-Lactobacilli bacterial or viral species in the vagino-cervical microbiome. *Bifidobacterium* was seen with older age at sexual debut or marriage; *Streptococcus agalactiae* was found in those who suffered a heavy menstrual flow. Other features such as mood fluctuations and facial speckles could potentially be deduced from the vagino-cervical microbiome. *Lactobacillus crispatus* and *Prevotella bivia* could be shared between the gut and the vagino-cervical microbiome. Oral microbiome, plasma vitamins, metals, amino acids and hormones showed clear associations with the vagino-cervical microbiome. Our results offer an unprecedented glimpse into the microbiota of the female reproductive tract and point to disease susceptibilities that may be relieved by behavioral changes.

Highlights:

1 Shotgun sequencing of 1148 vagino-cervical samples reveal rare subtypes;

Other omics such as moods and skin features can be predicted by the vagino-cervical microbiome;

Delivery mode, breast-feeding, and means of contraception make a lasting impact;

With dwindled Lactobacilli, postmenopausal samples harbor more viruses.

INTRODUCTION

The human body is a supra-organism consisting of more microbial cells than human cells. Studies in human cohorts and animal models have revealed an integral role played by the gut microbiota in metabolic and immunological functions. Disease markers as well as prebiotic or probiotic interventions are being actively developed. Studies on other mucosal sites such as the vagina and the mouth lag behind; the potential for the microbiota in these body sites has not been fully realized. The presence of microorganisms beyond the cervix (i.e. the upper reproductive tract) is increasingly recognized even in non-infectious conditions (Aagaard et al., 2014; Chen et al., 2017; Li et al., 2018), with much debated implications for women's and infants' health.

Lactobacilli have long been regarded as the keystone species of the vaginal microbiota. Lactic acid produced by these microorganisms helps maintain a low vaginal pH of 3.5-4.5, and wards off pathogenic micro-organisms. Prevention of the human immunodeficiency virus (HIV) and other sexually transmitted infections (STIs), preterm birth and bacterial vaginosis (BV) has been a major effort. Due to the over 90% of human sequences in female reproductive tract samples (Li et al., 2018; Methé et al., 2012), most studies of the vaginal microbiota used 16S rRNA gene amplicon

sequencing, which lacked a view of the overall microorganism communities including bacteria, archaea, viruses and fungi, as well as the functional capacity encoded. Besides infection, studies on current sexual activity and the menstrual cycle occurred naturally to the vaginal microbiota field (Gajer et al., 2012; Ravel et al., 2010). However, lasting impacts from other potentially important factors such as sexual debut, pregnancy and breast-feeding have not been looked at in a reasonably large cohort.

As a sizable reservoir of microbes instead of a transient entity, the female reproductive tract microbiota might also reflect conditions in other body sites. It is however not clear whether other omics in circulation or in the intestine cross-talks with the vagino-cervical microbiome. Intersecting with hormones, metabolic and immune functions, we find it intriguing to explore the potential link of the vagino-cervical microbiome to the brain and the face.

Here, we report the full spectrum of the vagino-cervical microbiome from 1148 healthy women using metagenomic shotgun sequencing. We also comprehensively measured parameters such as fecal/oral microbiome, plasma metabolites, medical test data, immune indices, physical fitness test, facial skin imaging, as well as female life history questionnaire, lifestyle questionnaire, and psychological questionnaire (Figure 1). Taken together, our work pinpoints other metadata or omics that can predict or be predicted from the microbiota in the female reproductive tract, which would illuminate future designs of population cohorts as well as 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 aged 21-52 (median 29, 95% CI: 23–39) were recruited during a physical examination as the initial study cohort (Figure 1; Table S1). 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 (Figure 2).

In agreement with 16S rRNA gene amplicon sequencing data from the US (Ravel et al., 2010), the vagino-cervical microbiota of this Asian cohort was mostly Lactobacilli-dominated, while 19.19% of women harbored over 50% *Gardnerella vaginalis* (Figure 2, Figure S1A). Some of the volunteers were dominated by *Atopobium vaginae* but not *G. vaginalis*, which were commonly believed to co-occur in bacterial vaginosis (BV) (Fredricks et al., 2005). 22.67% of the cohort was dominated by *Lactobacillus iners* (Figure 2, Figure S1A), 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 (Figure 2, Figure S1A). Other microorganisms including *Clamydia trachomatis*, *Ureaplasma parvum*, human papillomavirus (HPV), herpesviruses, *Haemophilus influenzae* and Enterobacteria phage were abundant in some individuals (Figure 2). According to the metagenomic

data, the vagino-cervical microbiome varied widely among individuals (Figure 2, Figure S1A), 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 (Figure S1B), suggesting that in future studies lower amounts of sequencing could be used for these compared to other types.

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, current breast-feeding, contraceptive methods, menstrual cycle, and sexual intercourse frequency (Figure 3).

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 (Figure 3).

As a whole, the vagino-cervical microbial composition showed the greatest explained variances for these questionnaire data collected for the female reproductive tract samples , followed by other data collected on the same day, such as fecal microbiome composition, psychological questionnaire, plasma metabolites, immune indices, facial skin imaging and medical test data (Figure 4A). Self-assessed depression was the strongest factor in the psychological questionnaire to be predicted by vagino-cervical

microbiome composition, spots on the cheeks constituted the best predicted factor in facial skin imaging, and plasma phosphoserine was ranked the first metabolites (Figure 4B).

Specific influences from pregnancy histories, contraception and menstrual symptoms

Marriage was one of the most significant factors to influence the vagino-cervical microbiome (Figure 3). 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* (Figure 5A**,** Table S2). Compared to unmarried women, married women had higher concentrations of plasma vitamin D, and their plasma testosterone (responsible for sexual drive), dehydroepiandrosterone (DHEA), and creatinine had declined (Figure 5A). The age at vaginal sexual debut showed positive correlations with *Gardnerella vaginalis* and *Scardovia* species, and negative associations with *L. crispatus* and *L. iners* (Figure 5B**,** Table S2).

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 (Figure 5C**,** Table S2, Figure S2). We observed increased concentrations of plasma vitamin D, blood glucose and direct bilirubin 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*, *L. jensenii*, *Bartonella* species, *Prevotella timonensis*, and increased *Lactobacillus* sp. 7_1_47FAA (Figure 5D, Table S2). We observed a lower relative abundance of *Ureaplasma parvum* in individuals who delivered by cesarean section (Figure 5E, Table S2), a bacterium commonly isolated from pregnant women (Anderson et al., 2013) and recently reported in the lower respiratory tract of preterm infants (Pattaroni et al., 2018). Compared with vaginal deliveries, the plasma concentrations of testosterone, androstenedione, methionine, threonine and tryptophan were significantly lower in women who experienced caesarean section (Figure 5D, 5E). In addition, these women more often suffered from an abnormal leucorrhoea (Figure 5E).

In 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 (Figure 5F). Alkaline phosphatase, plasma vitamin D and lead (Pb) were found to be more abundant in the breast-feeding individuals, consistent with slight reduction of progesterone and creatinine levels in lactating women (Figure 5F). 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 (Figure 5C-5F).

Integrated association network using a wisdom of crowds approach (Marbach et al., 2012) also revealed interesting patterns (Spearman's correlation, random forest and linear regression, Figure 6). The relative abundance of *G. vaginalis* was negatively

associated with sleep and appetite, and mental state, while positively associated with plasma levels of mercury (Hg) and vitamin E (Figure 6). *L. iners* negatively correlated with times of spontaneous abortion, days since last menstrual bleeding, condom usage and the plasma concentration of vitamin D and vitamin A, and positively correlated with plasma concentrations of hemoglobin, asparagine and alanine. In addition, vaccination history, such as those against HPV may be associated with higher relative abundance of *L. iners* (Figure 6). Vaginal douching within the past month positively correlated with *E. coli* in the gut microbiome (Figure 6), which might contribute to urinary tract infections.

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 (Table S2, Figure 6). Oral contraceptives, on the other hand, were associated with increased *Ureaplasma parvum*, *Comamonas* species and Mycoplasmataceae (Figure 5G, Table S2, and Figure 6). *Comamonas* has been implicated in fecundity in *C. elegans* and identified as a marker for infertility due to endometriosis (Chen et al., 2017; MacNeil et al., 2013). Moreover, oral contraceptives exhibited a positive correlation with plasma homocitrulline, and as yet not significant declines in mental health and butyrate-producing fecal microbiota species such as *Faecalibacterium prausnizi* and *Roseburia intestinalis* (Figure 5G), which could potentially contribute to the known side effect of oral contraceptives in cardiovascular risk. Thus, although lack of contraception could lead to infection, none of the contraceptive methods appeared healthy for the microbiome in the long run.

Menstrual phases are known to influence the microbiota in the female reproductive tract (Chen et al., 2017; Gajer et al., 2012). 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 (Figure 5H). 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 (Figure 6). *Streptococcus agalactiae*, *Propionibacterium acnes* and *Dialister micraerophilus* were positively associated with a heavier flow (Figure 6). Many women experience dysmenorrhea during menses. Individuals with dysmenorrhea were enriched for Pseudomonadales, *Acinetobacter* and *Moraxellaceae*, while lower in plasma level of histidine (Figure 5I), consistent with these bacteria producing histidine decarboxylases to convert histidine into histamine (Cundell et al., 1991).

Omics factors associated with vaginal types, immune and metabolic functions

The omics data also significantly differed among individuals with the different vaginal types (Figure S1, Figure S3). The women with types of *B. breve* and *G. vaginalis* had older age at vaginal sexual debut (Figure S3A), and the results advocate for getting a partner at a younger age. The types characterized by *L. crispatus* and *L. jensenii* were overrepresented in the women who were not married yet or had fewer pregnancies (Figure S3B, C). We observed higher plasma levels of proline and creatinine in the *S. anginosus*-type individuals, but lower plasma levels of 11-deoxycortisol and 17α-

hydroxyprogesterone (Figure S3D-G). The plasma concentrations of 11-deoxycortisol, 17α-hydroxyprogesterone and testosterone were higher in individuals of the *L. crispatus*-type than in *L. iners* and *G. vaginalis* type individuals (Figure S3F-H). Vaginal types also appeared associated with facial skin feature such as *A. vaginae* and *L. johnsonii* types had higher rates of facial skin problem than other types (Figure S3I-L). Fecal *Neisseria* species, *Haemophilus* species and *Aggregatibacter* species showed greater abundance in the *L. jensenii*-type (Figure S3M-O).

While lacking significant association 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, synthesis of branched-chain amino acids, histidine and arginine (Figure S4). Red blood cell counts, plasma vitamin A and plasma hydroxyl vitamin D levels negatively associated with CDP-diacylglycerol biosynthesis pathways, consistent with presence of diacylglycerol kinase in Lactobacilli with anti-inflammatory functions (Andrada et al., 2017; Ganesh et al., 2017). Fecal *Coprococcus comes*, a bacterium previously reported to associate with cytokine response to *Candida albicans* (Schirmer et al., 2016), was seen here to associate with isoleucine pathways in the vagino-cervical microbiome (Figure S4).

Validation in a more heterogeneous cohort along with oral microbiome data

In an independent cohort of 632 individuals that differed in age distribution (Figure 1, Table S3), we again investigated the factors that explain the vagino-cervical microbiome

and vice versa. Consistent with the initial cohort of 516 individuals, questionnaire entries such as pregnancy history, breast-feeding, delivery mode and times were found as the most important factors to influence the vagino-cervical microbiome in this cohort (Figure S5). The oral microbiome, which was reported to show similarity to the placenta microbiome (Aagaard et al., 2014), showed notable associations with the vagino-cervical microbiome (Table S4). Oral Human herpesvirus was positively correlated with Bifidobacteriaceae and *G. vaginalis* in vagina. *Scardovia wiggsiae*, a bacterium previously reported to associate with early childhood caries(Kressirer et al., 2017), showed positive correlation with vaginal *Staphylococcus* species. Oral *Treponema lecithinolyticum* showed positive correlation with *Dialister micraerophilus*. Besides, we observed fecal *L. crispatus* to associate with the same species in the vagino-cervical sample.

45 of the volunteers were postmenopausal, an age group untouched by HMP (median age 54, 95% CI: 52-55). Metagenomic shotgun data clearly showed diminished Lactobacilli in their vagino-cervical microbiome, while the mean proportion of viral sequences reached 37% (Figure S6). HPV was not the most abundant or prevalent virus in these postmenopausal individuals; a diverse range of viruses named after plant or animal hosts could be detected. In contrast, there was no sign of fungal boom. The non-Lactobacilli bacterial species were also known as vaginal or oral species, with overgrowth of *Escherichia* in very few individuals (Figure S6A).

The integrated association network between omics also revealed patterns that were highly similar to the initial cohort (Figure 7, Table S5). Age, pregnancy history, current breast-feeding showed negative correlations with relative abundances of *L. crispatus*, *L. jensenii*, *L. vaginalis*, *L. iners* and *U. parvum*. We also reconfirmed a lower relative abundance of *U. parvum* and a higher abundance of *P. timonensis* and *P. buccalis* in individuals who delivered by cesarean section. The increased concentrations of androstenedione, dehydroepiandrosterone, testosterone, 11-deoxycorticosterone and progesterone and decreased concentration of Lead (Pb) were observed in women with higher abundance of *L. crispatus*. We observed the increased concentrations of threonine, arginine, serine, and alanine and decreased concentrations of 17α-hydroxyprogesterone, progesterone, 25-hydroxy vitamin D, carnosine, and alanine aminotransferase in women with higher abundance of *U. parvum*. Serum creatinine showed positive correlations with *L. crispatus* and *L. vaginalis*, and negative associations with *A. vaginae*, *Dialister micraerophilus*, *Finegoldia magna*, *Streptococcus agalactiae*, *P. bivia*, *L. gasseri* and *P. timonensis*. The BV species *P. bivia* was negatively associated with fecal *Butyricimonas* species and *Clostridia* species, while positively associated with fecal *P. bivia*, suggesting a fecal reservoir for recurrent vaginal infections. The wrinkles and red area in forehead were relieved in individuals who enriched for *L. iners*. Together, these results confirmed the long-range links between the vagino-cervical microbiome and other body sites.

DISCUSSION

As the largest metagenomic study for the vagino-cervical microbiome, our data revealed lesser known subtypes for the vaginal microbiota, and put into context viral and fungal sequences present in healthy women. Our multi-omic data on a cohort of young women 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 the human immunodeficiency virus (HIV) 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. The vagino-cervical microbiome of postmenopausal women was also probed for the first time, revealing a myriad of viral and bacterial species. While fungal growth might be unfavorable both due to high pH and lack of glycans.

By reaching to the cervix with a cytobrush, reflecting both the upper and the lower reproductive tract (Chen et al., 2017; Li et al., 2018), we show here that the association with the fecal microbiome was overally weak, but involved both healthy (*L. crispatus*) and unhealthy (*P. bivia*) species. The plasma metabolites and T cell receptor types associated with the vagino-cervical microbiome are distinct from those associated with the fecal microbiome (Jie et al., 2019). The vagino-cervical microbiome also better predicted facial skin features compared to the fecal microbiome (Jie et al., 2019). 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. Beauty industry might be attracted to this important ecosystem in order to radiant from within.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

KEY RESOURCE TABLE

CONTACT FOR REAGENT AND RESOURCE SHARING

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Main Cohort

Validation Cohort

METHODS DETAILS

Demographic Data Collection

Samples Collection

DNA extraction and metagenomics shotgun sequencing

Amino Acid Measurements

Hormone Measurements

Trace element Measurements

Water-soluble Vitamins Measurements

Fat-soluble Vitamins Measurements

Immune indices Measurements

Medical Parameters

Facial Skin feature

Physical fitness test

QUANTIFICATION AND STATISTICAL ANALYSIS

Quality control, Taxonomic annotation and abundance calculation

Random-forest on the influence of female life history factors

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

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

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

Association between microbiome pathways and other omics

DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and five tables and can be found with this article online at XXX.

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., Z.J. and X.Z. wrote the manuscript. All authors contributed to data and texts in this manuscript.

DECLARATION OF INTERESTS

The authors declare no competing financial interest.

REFERENCES

Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The placenta harbors a unique microbiome. Sci. Transl. Med. *6*, 237ra65.

Anderson, B.L., Mendez-Figueroa, H., Dahlke, J.D., Raker, C., Hillier, S.L., and Cu-Uvin, S. (2013). Pregnancy-induced changes in immune protection of the genital tract: defining normal. Am. J. Obstet. Gynecol. *208*, 321.e1-9.

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

Chen, C., Song, X., Wei, W., Zhong, H., Dai, J., Lan, Z., Li, F., Yu, X., Feng, Q., Wang, Z., et al. (2017). The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. Nat. Commun. *8*, 875.

Cundell, D.R., Devalia, J.L., Wilks, M., Tabaqchali, S., and Davies, R.J. (1991). Histidine decarboxylases from bacteria that colonise the human respiratory tract. J. Med. Microbiol. *35*, 363–366.

Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H., Lu, H., Luber, J.M., Xia, M., Li, W., et al. (2018). Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. 1–8.

Fredricks, D.N., Fiedler, T.L., and Marrazzo, J.M. (2005). Molecular Identification of Bacteria Associated with Bacterial Vaginosis. N. Engl. J. Med. *353*, 1899–1911.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. *33*, 1–22.

Gajer, P., Brotman, R.M., Bai, G., Sakamoto, J., Schutte, U.M.E., Zhong, X., Koenig, S.S.K., Fu, L., Ma, Z., Zhou, X., et al. (2012). Temporal Dynamics of the Human Vaginal Microbiota. Sci. Transl. Med. *4*, 132ra52-132ra52.

Ganesh, B.P., Hall, A., Ayyaswamy, S., Nelson, J.W., Fultz, R., Major, A., Haag, A., Esparza, M., Lugo, M., Venable, S., et al. (2017). Diacylglycerol kinase synthesized by commensal Lactobacillus reuteri diminishes protein kinase C phosphorylation and histamine-mediated signaling in the mammalian intestinal epithelium. Mucosal Immunol.

Jie, Z., Liang, S., Ding, Q., Tang, S., Wang, D., Zhong, H., Jia, H., and Xu, X. (2019). A multi-omic cohort as a starting point for a healthy gut microbiome. To Be Submitt. Soon.

Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., and Huttenhower, C. (2018). Host genetic variation and its microbiome interactions within the Human Microbiome Project. Genome Med. *10*, 6.

Kressirer, C.A., Smith, D.J., King, W.F., Dobeck, J.M., Starr, J.R., and Tanner, A.C.R. (2017). Scardovia wiggsiae and its potential role as a caries pathogen. J. Oral Biosci. *59*, 135–141.

Li, F., Chen, C., Wei, W., Wang, Z., Dai, J., Hao, L., Song, L., Zhang, X., Zeng, L., Du, H., et al. (2018). The metagenome of the female upper reproductive tract. Gigascience *7*.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Prifti, E., Nielsen, T., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. Nat. Biotechnol. *32*, 834–841.

Liaw, A., and Wiener, M. (2001). Classification and Regression by RandomForest.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In Advances in Neural Information Processing Systems 26, C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, eds. (Curran Associates, Inc.), pp. 431–439.

MacNeil, L.T., Watson, E., Arda, H.E., Zhu, L.J., and Walhout, A.J.M. (2013). Diet-induced developmental acceleration independent of TOR and insulin in C. elegans. Cell *153*, 240–252.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., DREAM5 Consortium, Kellis, M., Collins, J.J., et al. (2012). Wisdom of crowds for robust gene network inference. Nat. Methods *9*, 796–804.

Methé, B.A., Nelson, K.E., Pop, M., Creasy, H.H., Giglio, M.G., Huttenhower, C., Gevers, D., Petrosino, J.F., Abubucker, S., Badger, J.H., et al. (2012). A framework for human microbiome research. Nature *486*, 215–221.

Moll1, J.M., Rausch, P., Eriksen, C., Nielsen, H.B., Feng, Q., Xia, H., and Brix, S. (2019). Pathway-based approach to classify unannotated gut bacteria associated with insulin sensitivity. Under Revis. Nat Microbiol.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol. *32*, 822–828.

Pattaroni, C., Watzenboeck, M.L., Schneidegger, S., Kieser, S., Wong, N.C., Bernasconi, E., Pernot, J., Mercier, L., Knapp, S., Nicod, L.P., et al. (2018). Early-Life Formation of the Microbial and Immunological Environment of the Human Airways. Cell Host Microbe *24*, 857–865.e4.

Qin, J., Li, Y., Cai, Z., Li, S.S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S.K., Mcculle, S.L., Ault, K., Peralta, L., and Forney, L.J. (2010). Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. *108*, 4680–4687.

Schirmer, M., Smeekens, S.P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E.A., Jansen, T., Jacobs, L., Bonder, M.J., Kurilshikov, A., et al. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. Cell *167*, 1125–1136.e8.

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

Shannon, P.T., Grimes, M., Kutlu, B., Bot, J.J., and Galas, D.J. (2013). RCytoscape: tools for exploratory network analysis. BMC Bioinformatics *14*, 217.

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

Wickham, H. (2015). ggplot2. Wiley Interdiscip. Rev. Comput. Stat. *3*, 180–185.

Zhang, W., Du, Y., Su, Z., Wang, C., Zeng, X., Zhang, R., Hong, X., Nie, C., Wu, J., Cao, H., et al. (2015). IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. Genetics *201*, 459–472.

FIGURE LEGENDS

Figure 1. Schematic overview of the study.

The study consists two cohorts with varying characteristics. 516 volunteers recruited in 2017 are regarded as the initial study cohort, and 632 volunteers recruited in 2018 are regarded as the validation cohort. Vagino-cervical microbiome, fecal microbiome, metabolites, medical test data, physical fitness test data, facial skin feature and three kinds of questionnaires are collected in both cohorts. Immune indices are only shown in the initial study cohort, and saliva microbiome is only shown in the validation cohort.

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

The microbial composition in each sample at the species level according to MetaPhylAn2 is shown. The samples are ordered according to hierarchical clustering. Taxa names in black and blue denote bacteria and viruses, respectively.

Figure 3. Factors from 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 5-fold cross-validated random forest (RFCV) importance on the microbiome composition.

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

(A) 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 vagino-cervical microbiome to omics data; numbers on the curved arrows indicate adjusted R-squared from omics data to vagino-cervical microbiome . (B) Top 8 factors in each type of omics that are predicted by the vagino-cervical microbiome. The length of the bars represents RFCV Rcorrelation coefficient between the actual measurements and the values predicted by 5-fold cross-validated random forest models.

Figure 5. Specific influences from reproductive factors on omics data in the initial study cohort.

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 **Figure 3A**. The metabolites such as amino acids, hormones and vitamins are measured in plasma, and trace elements are measured in whole blood. The blood biochemistry such as alkaline phosphatase, fasting blood sugar, direct bilirubin, creatinine, total bilirubin, high-density lipoprotein is measured in serum. The length of the bars represents Spearman's correlation coefficient between the factor and the omics data. A plus sign denotes BH-adjusted P-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.

Figure 6. Wisdom of the crowds for the association network between vagino-cervical microbial species and other omics data in the initial study cohort.

Results from generalized linear model with penalty (cv. glmnet), random forest (RFCV) and Spearman's correlation are integrated and then visualized in CytoScape. Red lines, negative associations; cyan lines, positive associations.

Figure 7. Wisdom of the crowds for the association network between vagino-cervical microbial species and other omics data in the initial study cohort and the validation cohort.

Results from generalized linear model with penalty (cv. glmnet), random forest (RFCV) and Spearman's correlation are integrated and then visualized in CytoScape. The correlation is shown when the same enrichment direction in both cohorts and BH-adjusted P-value <0.1. Red lines, negative associations; cyan lines, positive associations.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Cervical samples	This paper	N/A
Fecal samples	This paper	N/A
Saliva samples	This paper	N/A
Blood samples	This paper	N/A
Critical Commercial Assays		
FTA card	BGI	N/A
MGIEasy sample collection kit	MGI	http://www.mgitech.cn/article/detail/mgieasyfbybcjtzjz111.html
MGIEasy DNA Library Prep Kit	MGI	http://www.mgitech.cn/article/detail/MGIEasy%20DNA.html
Deposited Data		
Raw sequencing data	This paper	CNGB database accession CNP0000287
Software and Algorithms		
Metaphlan2	(Segata, 2012)	https://bitbucket.org/biobakery/metaphlan2
R	R Development Core Team	https://www.r-project.org/
randomForest R package	(Liaw and Wiener, 2001)	https://CRAN.R-project.org/package=randomForest
packfor R package	N/A	https://rdrr.io/rforge/packfor/
CytoScape	(Shannon et al., 2013)	https://cytoscape.org/
glmnet R package	(Friedman et al., 2010)	https://CRAN.R-project.org/package=glmnet
ggplot2 R package	(Wickham, 2015)	https://CRAN.R-project.org/package=ggplot2
HUMANn2	(Kolde et al., 2018)	https://bitbucket.org/biobakery/humann2/downloads/
pheatmap R package	N/A	https://CRAN.R-project.org/package=pheatmap

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to, and will fulfilled by, the Lead Contact, Huijue Jia (jiahuijue@genomics.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Initial Study Cohort

All the 516 volunteers were recruited between May 2017 and July 2017 during annual physical examination in BGI-Shenzhen, China. Baseline characteristics of the cohort are shown in Table S1.

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

Validation Cohort

An independent cohort of 632 were also recruited between May 2018 and July 2018 during annual physical examination in BGI-Shenzhen, China. The volunteers who underwent physical examination for two consecutive years were excluded from the validation cohort. The process of omics data collection was consistent with the initial study cohort. In addition, saliva samples were collected only in this cohort. Baseline characteristics of the cohort are shown in Table S2.

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

METHODS DETAILS

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. The lifestyle questionnaire contained age, disease history, eating and exercise habits. The psychological questionnaire contained the evaluation of irritability, dizziness, frustration, fear, appetite, self-confidence, resilience (Table S1, S3).

Samples Collection

Cervical samples were collected and smeared in the Flinders Technology Associates (FTA) cards by the doctor during gynecological examination. Fecal samples and saliva

samples were self-collected by volunteers. Cervical samples, fecal samples and saliva 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(Chen et al., 2017)(Qin et al., 2012), respectively. Metagenomic sequencing was done on the BGISEQ-500 platform (50bp of singled-end reads for cervical samples collected in the initial study cohort, and four libraries were constructed for each lane; 100bp of singled-end reads for fecal samples collected in the initial study cohort, and four libraries were constructed for each lane; 100bp of paired-end reads for cervical samples, fecal samples and saliva sample collected in the validation cohort, and eight libraries were constructed for each lane)(Fang et al., 2018).

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.

Trace element Measurements

200 µl of whole blood were 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_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 were 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 were 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 were dried by nitrogen flow. 80 µl 80% acetonitrile were added to the residues, 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 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 were 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 singled-end reads. The data processing was performed using Immune IMonitor(Zhang et al., 2015).

Medical Parameters

All the volunteers were recruited during the physical examination. The medical test including blood tests, urinalysis, routine examination of cervical secretion. All the medical parameters were measured by the physical examination center and shown in Table S1and Table S3.

Facial Skin feature

The volunteers were required to clean their face without makeup after they got up in the morning. The volunteer's frontal face was photographed by VISIA-CRTM imaging system (Canfield Scientific, Fairfield, NJ, USA) 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, S3). The percentile of index was calculated based on the index value ranked in the age-matched database (Table S1, S3). 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 (Table S1, S3). 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 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.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quality control, Taxonomic annotation and abundance calculation

The sequencing reads of fecal samples and saliva samples were quality-controlled and then aligned to hg19 to remove human reads as described previously(Fang et al., 2018). Stringent condition for removal of host sequences was used for cervical samples(Li et al., 2018), through alignment to the GRCh38 reference.

Taxonomic assignment of the high-quality cervical metagenomic data and saliva metagenomic data were performed using MetaPhlAn2(Segata, 2012).

Taxonomic assignment of the high-quality fecal metagenomic data was performed using the reference gene catalog comprising 9,879,896 genes(Li et al., 2014). Taxonomy of the fecal MLGs/MGSs were then determined from their constituent genes, as previously described(Moll1 et al., 2019; Nielsen et al., 2014; Qin et al., 2012; Wang and Jia, 2016).

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 (Marbach et al., 2012) 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 details are 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(Louppe et al., 2013). 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 (Figure 5). CytoScape was also used to visualized the omics network (Figure 6). The validation cohort was analyzed using the same statistical method. Combining P-value is computed using Edgington method from metap package in R. Benjamini and Hochberg methods were used to adjusted the multiple test P-value. The correlation is identified based on BH-adjusted P-value <0.1 when the similar microbial distribution pattern shown in the initial study cohort and validation cohort.

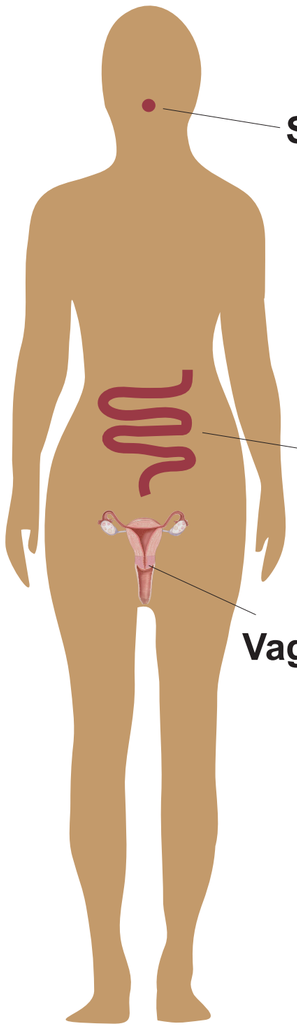
Association between microbiome pathways and other omics

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.

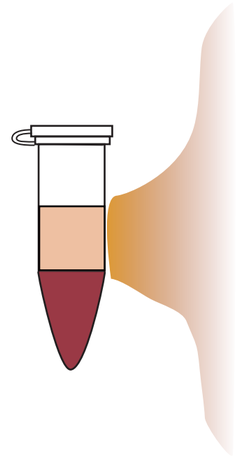
DATA AND SOFTWARE AVAILABILITY

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

1148 Healthy women (two cohorts)



Shotgun metagenomics



39 amino acides (plasma)

13 hormones (plasma)

9 vitamins (plasma)

9 trace elements (whole blood)

622 immune indices (buffy coat)

Blood biochemistry: alkaline phosphatase, fasting blood sugar, bilirubin, globulin, HDL, LDL, creatinine (serum)

BMI, chest circumference, uric acid, leucorrhea cleanliness

Metabolites

Immune indices

Medical test

Physical fitness test

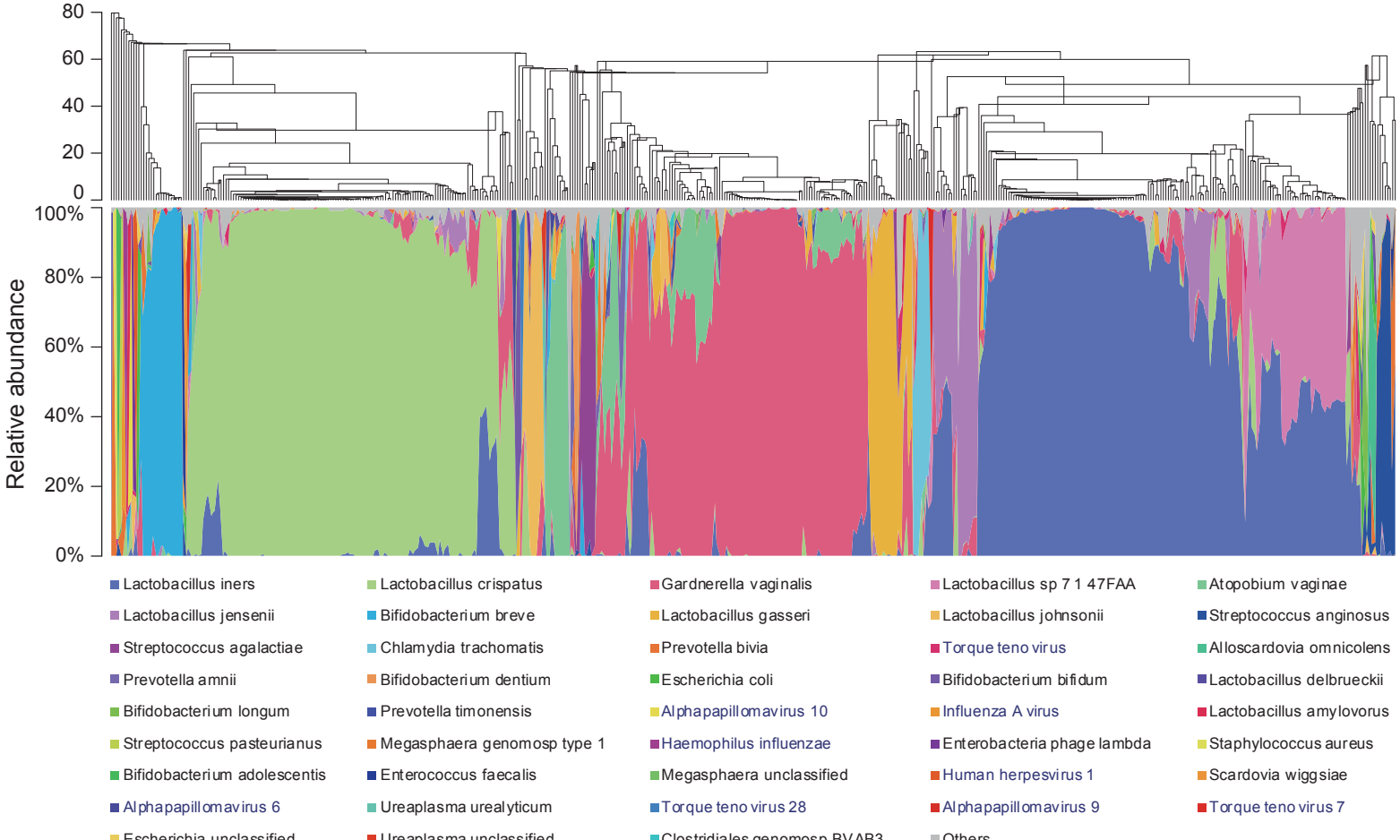
vital capacity
grip strength
sit and reach
sit-ups
step index
vertical jump
choice reaction time
single-legged stance with eyes closed

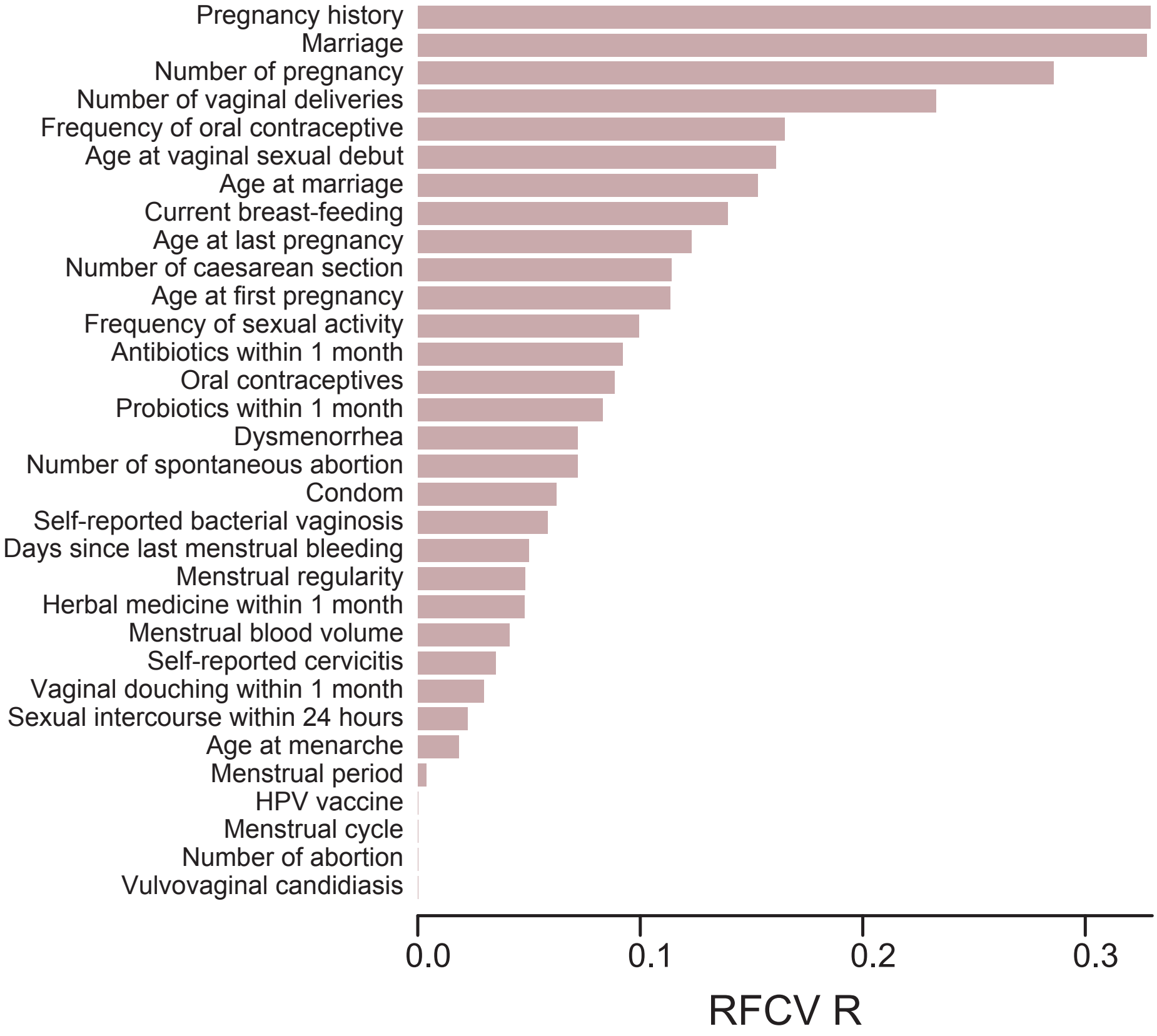
Facial skin feature

spots
pores
wrinkles
texture
UV spots
porphyrins
brown spots
red area

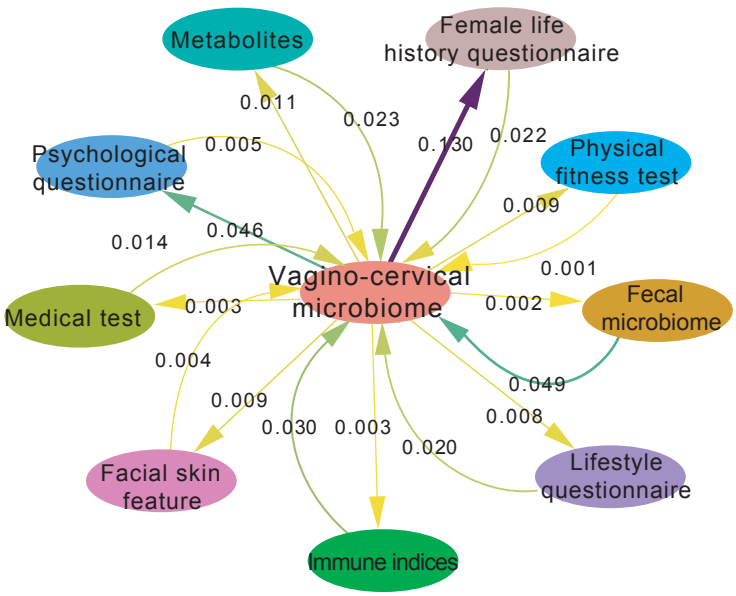
3 kinds of questionnaires

Female life history questionnaire
Lifestyle questionnaire
Psychological questionnaire

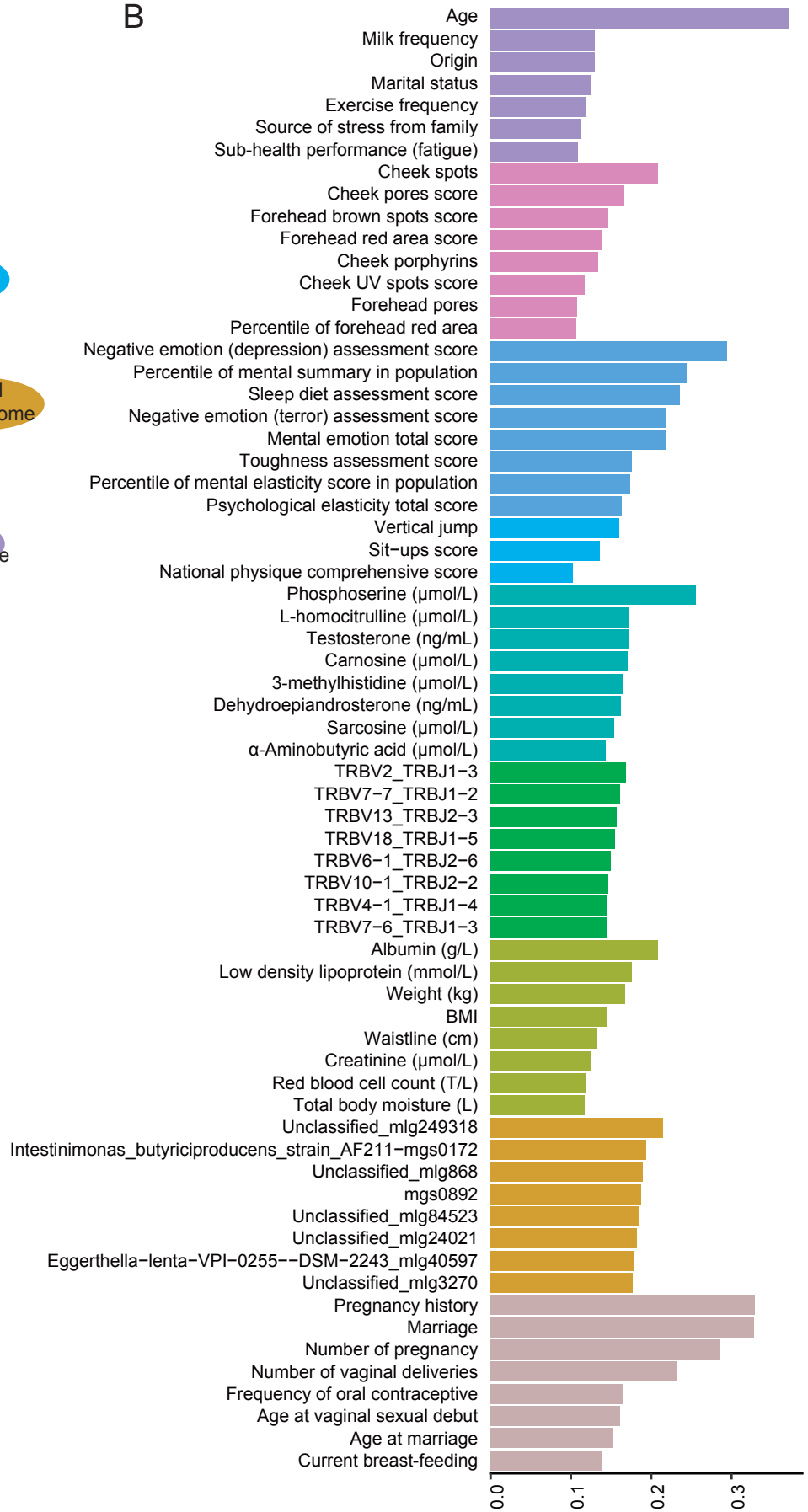

80
60
40
20
0
100%
80%
60%
40%
20%
0%
Relative abundance
Lactobacillus iners
Lactobacillus crispatus
Gardnerella vaginalis
Lactobacillus sp 7 1 47FAA
Atopobium vaginae
Lactobacillus jensenii
Bifidobacterium breve
Lactobacillus gasseri
Lactobacillus johnsonii
Streptococcus anginosus
Streptococcus agalactiae
Chlamydia trachomatis
Prevotella bivia
Torque teno virus
Alloscardovia omnicolens
Prevotella amnii
Bifidobacterium dentium
Escherichia coli
Bifidobacterium bifidum
Lactobacillus delbrueckii
Bifidobacterium longum
Prevotella timonensis
Alphapapillomavirus 10
Influenza A virus
Lactobacillus amylovorus
Streptococcus pasteurianus
Megasphaera genomosp type 1
Haemophilus influenzae
Enterobacteria phage lambda
Staphylococcus aureus
Bifidobacterium adolescentis
Enterococcus faecalis
Megasphaera unclassified
Human herpesvirus 1
Scardovia wiggsiae
Alphapapillomavirus 6
Ureaplasma urealyticum
Torque teno virus 28
Alphapapillomavirus 9
Torque teno virus 7
Escherichia unclassified
Ureaplasma unclassified
Clostridiales genomosp BVAB3
Others


Pregnancy history
Marriage
Number of pregnancy
Number of vaginal deliveries
Frequency of oral contraceptive
Age at vaginal sexual debut
Age at marriage
Current breast-feeding
Age at last pregnancy
Number of caesarean section
Age at first pregnancy
Frequency of sexual activity
Antibiotics within 1 month
Oral contraceptives
Probiotics within 1 month
Dysmenorrhea
Number of spontaneous abortion
Condom
Self-reported bacterial vaginosis
Days since last menstrual bleeding
Menstrual regularity
Herbal medicine within 1 month
Menstrual blood volume
Self-reported cervicitis
Vaginal douching within 1 month
Sexual intercourse within 24 hours
Age at menarche
Menstrual period
HPV vaccine
Menstrual cycle
Number of abortion
Vulvovaginal candidiasis
0.0
0.1
0.2
0.3
RFCV R

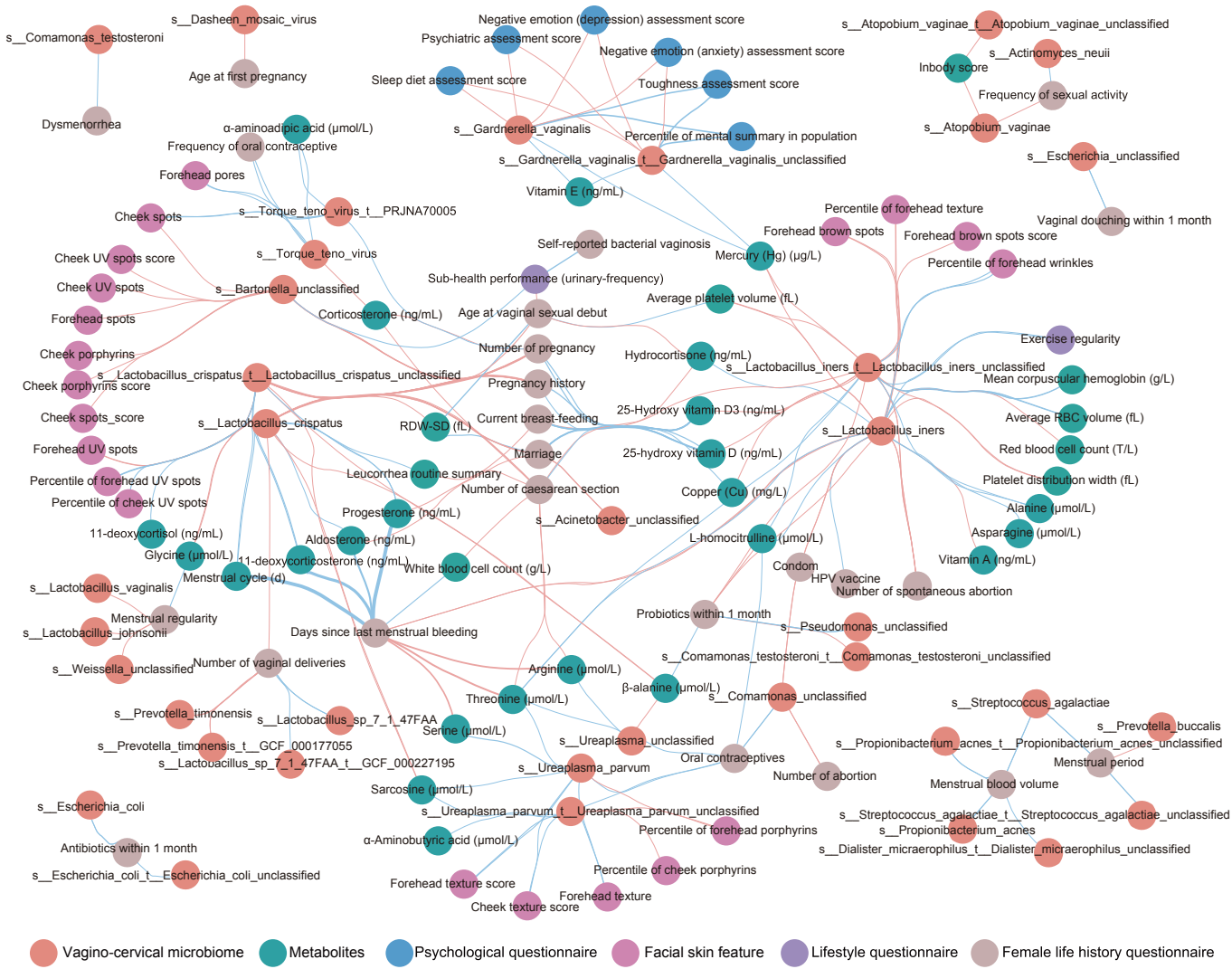
A

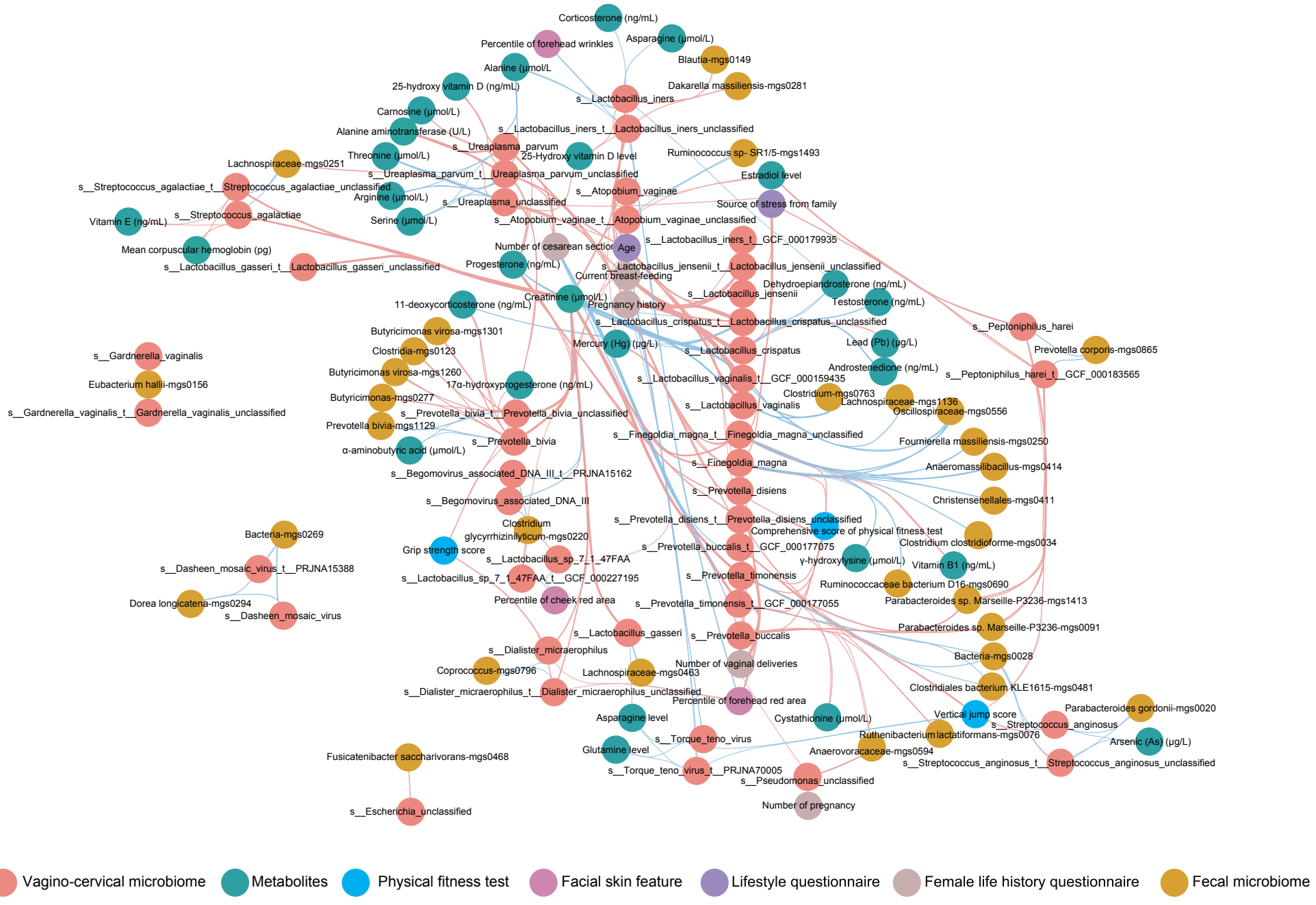

Metabolites
Female life history questionnaire
Psychological questionnaire
Physical fitness test
Vagino-cervical microbiome
Medical test
Fecal microbiome
Facial skin feature
Lifestyle questionnaire
Immune indices
0.011
0.023
0.130
0.022
0.005
0.009
0.014
0.046
0.001
0.003
0.002
0.004
0.049
0.009
0.008
0.030
0.003
0.020

B


Age
Milk frequency
Origin
Marital status
Exercise frequency
Source of stress from family
Sub-health performance (fatigue)
Cheek spots
Cheek pores score
Forehead brown spots score
Forehead red area score
Cheek porphyrins
Cheek UV spots score
Forehead pores
Percentile of forehead red area
Negative emotion (depression) assessment score
Percentile of mental summary in population
Sleep diet assessment score
Negative emotion (terror) assessment score
Mental emotion total score
Toughness assessment score
Percentile of mental elasticity score in population
Psychological elasticity total score
Vertical jump
Sit−ups score
National physique comprehensive score
Phosphoserine (μmol/L)
L-homocitrulline (μmol/L)
Testosterone (ng/mL)
Carnosine (μmol/L)
3-methylhistidine (μmol/L)
Dehydroepiandrosterone (ng/mL)
Sarcosine (μmol/L)
α-Aminobutyric acid (μmol/L)
TRBV2_TRBJ1−3
TRBV7−7_TRBJ1−2
TRBV13_TRBJ2−3
TRBV18_TRBJ1−5
TRBV6−1_TRBJ2−6
TRBV10−1_TRBJ2−2
TRBV4−1_TRBJ1−4
TRBV7−6_TRBJ1−3
Albumin (g/L)
Low density lipoprotein (mmol/L)
Weight (kg)
BMI
Waistline (cm)
Creatinine (μmol/L)
Red blood cell count (T/L)
Total body moisture (L)
Unclassified_mlg249318
Intestinimonas_butyriciproducens_strain_AF211−mgs0172
Unclassified_mlg868
mgs0892
Unclassified_mlg84523
Unclassified_mlg24021
Eggerthella−lenta−VPI−0255−−DSM−2243_mlg40597
Unclassified_mlg3270
Pregnancy history
Marriage
Number of pregnancy
Number of vaginal deliveries
Frequency of oral contraceptive
Age at vaginal sexual debut
Age at marriage
Current breast-feeding
0.0
0.1
0.2
0.3

A
Marriage
Age
Source of stress from family
Marital status
Source of stress from emotion
Cheek porphyrins
Cheek spots
Forehead porphyrins
Cheek UV spots
Forehead UV spots
Psychiatric assessment score
Choice reaction time score
Grip strength (kg)
Single-legged stance with eyes closed score
Sit and reach (cm)
25-hydroxy vitamin D (ng/mL)
25-hydroxy vitamin D3 (ng/mL)
Dehydroepiandrosterone (ng/mL)
Testosterone (ng/mL)
Waist−hip ratio
Waistline (cm)
Chest circumference (cm)
BMI
Creatinine (μmol/L)
Prevotella copri−mgs1172
Veillonella tobetsuensis−mgs1162
Weissella confusa−mgs0469
Holdemania massiliensis−mgs0292
Oscillibacter sp. KLE 1745−mgs0204
f__Lactobacillaceae
g__Lactobacillus
f__Bartonellaceae
s__Comamonas_testosteroni_t__Comamonas_testosteroni_unclassified
s__Comamonas_testosteroni
s__Acinetobacter_unclassified
g__Acinetobacter
f__Moraxellaceae
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
s__Lactobacillus_crispatus
−1 −0.5 0.0 0.5
B
Age at vaginal sexual debut
Age
Sub-health performance (urinary-frequency)
Drinking situation
Exercise strength
Cheek spots
Cheek pores
Cheek porphyrins score
Forehead wrinkles
Forehead brown spots score
Choice reaction time score
Ornithine (μmol/L)
Tetrahydrofolate (ng/mL)
Hydrocortisone (ng/mL)
Tryptophan (μmol/L)
RDW-CV (%)
RDW-SD (fL)
Average platelet volume (fL)
Lymphocytes (%)
Monocytes (%)
Oscillospiraceae−mgs0502
Lachnospiraceae−mgs0996
Roseburia_faecis−mgs0299
Alistipes_obesi−mgs0342
Bilophila_wadsworthia−mgs0061
p__Actinobacteria
c__Actinobacteria
o__Bifidobacteriales
f__Bifidobacteriaceae
s__Lactobacillus_crispatus
g__Lactobacillus
f__Lactobacillaceae
p__Firmicutes
o__Lactobacillales
c__Bacilli
−0.25 0.00 0.25 0.50
C
Pregnancy history
Age
Source of stress from family
Exercise regularity
Marital status
Cheek porphyrins
Percentile of cheek spots
Forehead porphyrins
Cheek porphyrins score
Percentile of cheek brown spots
Reaction time (s)
Sit-ups
25-Hydroxy vitamin D3 (ng/mL)
Dhydroepiandrosterone (ng/mL)
Methionine (μmol/L)
Androstenedione (ng/mL)
Testosterone (ng/mL)
Alkaline phosphatase (U/L)
Fasting blood sugar (mmol/L)
Direct bilirubin (μmol/L)
Waist−hip ratio
Creatinine (μmol/L)
Lachnospiraceae−mgs0164
Adlercreutzia_equolifaciens−mgs1330
Streptococcus_salivarius−mgs0906
Blautia−mgs1291
Blautia−mgs0130
c__Bacilli
p__Firmicutes
o__Lactobacillales
f__Lactobacillaceae
s__Lactobacillus_crispatus
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
g__Lactobacillus
s__Bartonella_unclassified
g__Bartonella
f__Bartonellaceae
−0.6 −0.3 0.0 0.3
D
Number of vaginal deliveries
Nap−habit
Cheek brown spots score
Percentile of forehead porphyrins
Percentile of cheek brown spots
Negative emotion (depression) assessment score
Compulsive symptom assessment score
Negative emotion (paranoid) assessment score
National physique comprehensive score
Single-legged stance with eyes closed (s)
17α-hydroxyprogesterone (ng/mL)
Menstrual cycle (d)
Progesterone (ng/mL)
11-deoxycorticosterone (ng/mL)
Cortisone (ng/mL)
Alkaline phosphatase (U/L)
Total bilirubin (μmol/L)
Indirect bilirubin (μmol/L)
Neutrophils (%)
Monocytes (%)
Lachnospiraceae−mgs0164
Lachnospiraceae−mgs0587
Oscillospiraceae−mgs0422
Clostridia−mgs1503
Coprobacillus_sp._3_3_56FAA−mgs0084
o__Viruses_noname
c__Viruses_noname
p__Firmicutes
f__Lactobacillaceae
s__Lactobacillus_crispatus
p__Proteobacteria
c__Alphaproteobacteria
o__Rhizobiales
s__Bartonella_unclassified
g__Bartonella
−0.2 0.0 0.2
E
Number of cesarean deliveries
Age
Fruit and vegetable intake per dose
Dietary structure
Source of stress from health
Marital status
Cheek porphyrins
Cheek UV spots
Forehead UV spots
Cheek spots
Cheek pores
Negative emotion (depression) assessment score
Compulsive symptom assessment score
Sit and reach (cm)
Tryptophan (μmol/L)
Threonine (μmol/L)
Methionine (μmol/L)
Androstenedione (ng/mL)
Testosterone (ng/mL)
Total neutrophils (g/L)
White blood cell count (g/L)
Waistline (cm)
Total lymphocyte count (g/L)
Leucorrhea routine summary
Haemophilus_parainfluenzae−mgs0101
Faecalibacterium_prausnitzii−mgs0432
Dorea_longicatena−mgs0294
Ruminococcus_lactaris−mgs0445
Blautia−mgs0130
o__Actinomycetales
p__Actinobacteria
c__Actinobacteria
c__Bacilli
p__Firmicutes
o__Lactobacillales
f__Lactobacillaceae
s__Lactobacillus_crispatus
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
g__Lactobacillus
−0.2 0.0 0.2
F
Current breast-feeding
Source of stress from family
Favor salty food
Working strength
Percentile of forehead porphyrins
Forehead texture
Forehead pores
Forehead texture score
Forehead pores score
Single-legged stance with eyes closed (s)
Single-legged stance with eyes closed score
Sit−ups score
Lead (Pb) (μg/L)
25-hydroxy vitamin D (ng/mL)
25-Hydroxy vitamin D3 (ng/mL)
Menstrual cycle (d)
Progesterone (ng/mL)
Alkaline phosphatase (U/L)
Hematocrit (L/L)
High-density lipoprotein (mmol/L)
Chest circumference (cm)
Creatinine (μmol/L)
Clostridia−mgs0062
Oscillibacter_sp._ER4−mgs1340
Oscillibacter_sp._ER4−mgs1040
Clostridia−mgs0123
Christensenellales−mgs0056
p__Actinobacteria
c__Actinobacteria
s__Lactobacillus_iners_t__Lactobacillus_iners_unclassified
c__Bacilli
p__Firmicutes
o__Lactobacillales
f__Lactobacillaceae
s__Lactobacillus_iners
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
g__Lactobacillus
−0.3 0.0 0.3 0.6 0.9
G
Oral contraceptives
Sub-health performance (leg-edema)
Sub-health performance (fatigue)
Favor oily food
Exercise frequency
Percentile of cheek texture
Forehead porphyrins
Cheek spots
Cheek pores
Cheek porescore
Negative emotion (terror) assessment score
Mental emotion total score
Negative emotion (anxiety) assessment score
Percentile of mental summary in population
Percentile of mental symptoms total score in population
Sit and reach score
Single-legged stance with eyes closed (s)
L-homocitrulline (μmol/L)
Taurine (μmol/L)
Mercury (Hg) (μg/L)
Valine (μmol/L)
Glutamic acid (μmol/L)
Waist−hip ratio
Waistline (cm)
Chest circumference (cm)
Body fat (kg)
Weight (kg)
Coprococcus_sp._HPP0074−mgs0232
Faecalibacterium_prausnitzii−mgs0328
Roseburia_intestinalis−mgs0158
Holdemania_massiliensis−mgs0292
Faecalibacterium_prausnitzii−mgs0238
s__Comamonas_unclassified
s__Ureaplasma_parvum_t__Ureaplasma_parvum_unclassified
s__Ureaplasma_parvum
g__Ureaplasma
f__Mycoplasmataceae
c__Mollicutes
p__Tenericutes
c__Bacilli
o__Lactobacillales
f__Lactobacillaceae
-0.2 0.0 0.2
H
Days since last menstrual bleeding
Favor sweet food
Fruit and vegetable intake per dose
Stay up late sleeps after 0:00 a.m.
Cheek brown spots score
Cheek red area score
Percentile of cheek pores
Cheek red area
Sit-ups
17α-hydroxyprogesterone (ng/mL)
Menstrual cycle (d)
Progesterone (ng/mL)
Threonine (μmol/L)
Arginine (μmol/L)
Monocytes (g/L)
Total neutrophils (g/L)
White blood cell count (g/L)
High-density lipoprotein (mmol/L)
Lymphocytes (%)
Bacteria−mgs0019
Oscillospiraceae−mgs0189
Eisenbergiella_tayi−mgs1385
Bifidobacterium_longum−mgs0625
Bacteroides_fragilis−mgs0079
o__Actinomycetales
p__Actinobacteria
c__Actinobacteria
g__Propionibacteriaceae_unclassified
f__Propionibacteriaceae
f__Actinomycetaceae
f__Comamonadaceae
p__Tenericutes
s__Lactobacillus_iners_t__Lactobacillus_iners_unclassified
s__Lactobacillus_iners
−0.4 0.0 0.4
I
Dysmenorrhea
High sugar and high fat food intake frequency
Favor hot food
Drinking situation
Mental stress
Favor plain food
Forehead porphyrins
Forehead spots
Cheek porphyrins score
Percentile of cheek brown spots
Percentile of forehead red area
Negative emotion (terror) assessment score
Mental emotion total score
Negative emotion (anxiety) assessment score
Negative emotion (interpersonal sensitivity) assessment score
Negative emotion (depression) assessment score
Comprehensive score of physical fitness test
Vitamin B5 (ng/mL)
Vitamin A (ng/mL)
Tetrahydrofolate (ng/mL)
Histidine (μmol/L)
Threonine (μmol/L)
Platelet count (g/L)
Total body moisture (L)
Red blood cell count (T/L)
BMI
Eubacterium_hallii−mgs0654
Christensenellales−mgs0411
Cloacibacillus_porcorum−mgs0124
Bacteroides−mgs1425
Ruminococcus_callidus−mgs0497
o__Pseudomonadales
g__Acinetobacter
f__Moraxellaceae
s__Propionibacterium_acnes_t__Propionibacterium_acnes_unclassified
s__Propionibacterium_acnes
g__Propionibacterium
s__Lactobacillus_iners_t__Lactobacillus_iners_unclassified
p__Firmicutes
s__Lactobacillus_crispatus
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
−0.1 0.0 0.1 0.2
Vagino-cervical microbiome
Fecal microbiome
Medical test
Metabolites
Physical fitness test
Psychological questionnaire
Facial skin feature
Lifestyle questionnaire


s__Dasheen_mosaic_virus
s__Comamonas_testosteroni
Age at first pregnancy
Dysmenorrhea
Negative emotion (depression) assessment score
Psychiatric assessment score
Negative emotion (anxiety) assessment score
Sleep diet assessment score
Toughness assessment score
s__Atopobium_vaginae_t__Atopobium_vaginae_unclassified
s__Actinomyces_neuii
Inbody score
Frequency of sexual activity
s__Atopobium_vaginae
α-aminoadipic acid (μmol/L)
Frequency of oral contraceptive
s__Gardnerella_vaginalis
Percentile of mental summary in population
s__Gardnerella_vaginalis_t__Gardnerella_vaginalis_unclassified
s__Escherichia_unclassified
Forehead pores
Vitamin E (ng/mL)
Cheek spots
s__Torque_teno_virus_t__PRJNA70005
Percentile of forehead texture
Vaginal douching within 1 month
Forehead brown spots
Forehead brown spots score
Self-reported bacterial vaginosis
Mercury (Hg) (μg/L)
Cheek UV spots score
s__Torque_teno_virus
Percentile of forehead wrinkles
Sub-health performance (urinary-frequency)
Cheek UV spots
s__Bartonella_unclassified
Average platelet volume (fL)
Forehead spots
Corticosterone (ng/mL)
Age at vaginal sexual debut
Exercise regularity
Number of pregnancy
Hydrocortisone (ng/mL)
Cheek porphyrins
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
s__Lactobacillus_iners_t__Lactobacillus_iners_unclassified
Mean corpuscular hemoglobin (g/L)
Cheek porphyrins score
Pregnancy history
Cheek spots_score
s__Lactobacillus_crispatus
RDW-SD (fL)
Current breast-feeding
25-Hydroxy vitamin D3 (ng/mL)
Average RBC volume (fL)
s__Lactobacillus_iners
Forehead UV spots
Marriage
25-hydroxy vitamin D (ng/mL)
Red blood cell count (T/L)
Percentile of forehead UV spots
Leucorrhea routine summary
Number of caesarean section
Copper (Cu) (mg/L)
Platelet distribution width (fL)
Percentile of cheek UV spots
Progesterone (ng/mL)
s__Acinetobacter_unclassified
Alanine (μmol/L)
11-deoxycortisol (ng/mL)
Aldosterone (ng/mL)
L-homocitrulline (μmol/L)
Asparagine (μmol/L)
Glycine (μmol/L)
11-deoxycorticosterone (ng/mL)
White blood cell count (g/L)
Condom
Vitamin A (ng/mL)
Menstrual cycle (d)
HPV vaccine
Number of spontaneous abortion
s__Lactobacillus_vaginalis
Menstrual regularity
Probiotics within 1 month
s__Pseudomonas_unclassified
s__Lactobacillus_johnsonii
Days since last menstrual bleeding
s__Comamonas_testosteroni_t__Comamonas_testosteroni_unclassified
s__Weissella_unclassified
Number of vaginal deliveries
Arginine (μmol/L)
β-alanine (μmol/L)
s__Comamonas_unclassified
Threonine (μmol/L)
s__Streptococcus_agalactiae
s__Prevotella_timonensis
s__Lactobacillus_sp_7_1_47FAA
s__Prevotella_buccalis
Serine (μmol/L)
s__Prevotella_timonensis_t__GCF_000177055
s__Ureaplasma_unclassified
s__Propionibacterium_acnes_t__Propionibacterium_acnes_unclassified
s__Lactobacillus_sp_7_1_47FAA_t__GCF_000227195
Oral contraceptives
Menstrual period
s__Ureaplasma_parvum
Number of abortion
Menstrual blood volume
Sarcosine (μmol/L)
s__Escherichia_coli
s__Ureaplasma_parvum_t__Ureaplasma_parvum_unclassified
s__Streptococcus_agalactiae_t__Streptococcus_agalactiae_unclassified
Percentile of forehead porphyrins
s__Propionibacterium_acnes
α-Aminobutyric acid (μmol/L)
Antibiotics within 1 month
s__Dialister_micraerophilus_t__Dialister_micraerophilus_unclassified
Percentile of cheek porphyrins
s__Escherichia_coli_t__Escherichia_coli_unclassified
Forehead texture score
Forehead texture
Cheek texture score
Vagino-cervical microbiome
Metabolites
Psychological questionnaire
Facial skin feature
Lifestyle questionnaire
Female life history questionnaire


Corticosterone (ng/mL)
Percentile of forehead wrinkles
Asparagine (μmol/L)
Blautia-mgs0149
Alanine (μmol/L
25-hydroxy vitamin D (ng/mL)
Dakarella massiliensis-mgs0281
s__Lactobacillus_iners
Carnosine (μmol/L)
Alanine aminotransferase (U/L)
s__Lactobacillus_iners_t__Lactobacillus_iners_unclassified
s__Ureaplasma_parvum
Threonine (μmol/L)
25-Hydroxy vitamin D level
Ruminococcus sp- SR1/5-mgs1493
Lachnospiraceae-mgs0251
s__Ureaplasma_parvum_t__Ureaplasma_parvum_unclassified
Estradiol level
s__Streptococcus_agalactiae_t__Streptococcus_agalactiae_unclassified
Arginine (μmol/L)
s__Atopobium_vaginae
s__Ureaplasma_unclassified
Source of stress from family
Vitamin E (ng/mL)
s__Streptococcus_agalactiae
Serine (μmol/L)
s__Atopobium_vaginae_t__Atopobium_vaginae_unclassified
Mean corpuscular hemoglobin (pg)
s__Lactobacillus_iners_t__GCF_000179935
Number of cesarean sectior
Age
s__Lactobacillus_gasseri_t__Lactobacillus_gasseri_unclassified
Progesterone (ng/mL)
s__Lactobacillus_jensenii_t__Lactobacillus_jensenii_unclassified
Current breast-feeding
Dehydroepiandrosterone (ng/mL)
s__Lactobacillus_jensenii
Creatinine (μmol/L)
Testosterone (ng/mL)
11-deoxycorticosterone (ng/mL)
Pregnancy history
s__Lactobacillus_crispatus_t__Lactobacillus_crispatus_unclassified
s__Peptoniphilus_harei
Butyricimonas virosa-mgs1301
Mercury (Hg) (μg/L)
Lead (Pb) (μg/L)
Prevotella corporis-mgs0865
s__Gardnerella_vaginalis
Clostridia-mgs0123
s__Lactobacillus_crispatus
Androstenedione (ng/mL)
s__Peptoniphilus_harei_t__GCF_000183565
Butyricimonas virosa-mgs1260
s__Lactobacillus_vaginalis_t__GCF_000159435
Eubacterium hallii-mgs0156
17α-hydroxyprogesterone (ng/mL)
Clostridium-mgs0763
Lachnospiraceae-mgs1136
Butyricimonas-mgs0277
s__Lactobacillus_vaginalis
Oscillospiraceae-mgs0556
s__Gardnerella_vaginalis_t__Gardnerella_vaginalis_unclassified
s__Prevotella_bivia_t__Prevotella_bivia_unclassified
Prevotella bivia-mgs1129
s__Finegoldia_magna_t__Finegoldia_magna_unclassified
Fournierella massiliensis-mgs0250
s__Prevotella_bivia
α-aminobutyric acid (μmol/L)
s__Finegoldia_magna
Anaeromassilibacillus-mgs0414
s__Begomovirus_associated_DNA_III_t__PRJNA15162
s__Prevotella_disiens
Christensenellales-mgs0411
s__Begomovirus_associated_DNA_III
s__Prevotella_disiens_t__Prevotella_disiens_unclassified
Clostridium
Comprehensive score of physical fitness test
Bacteria-mgs0269
glycyrrhizinilyticum-mgs0220
Clostridium clostridioforme-mgs0034
Grip strength score
s__Prevotella_buccalis_t__GCF_000177075
s__Lactobacillus_sp_7_1_47FAA
γ-hydroxylysine (μmol/L)
Vitamin B1 (ng/mL)
s__Dasheen_mosaic_virus_t__PRJNA15388
s__Lactobacillus_sp_7_1_47FAA_t__GCF_000227195
s__Prevotella_timonensis
Ruminococcaceae bacterium D16-mgs0690
Dorea longicatena-mgs0294
Percentile of cheek red area
s__Prevotella_timonensis_t__GCF_000177055
Parabacteroides sp. Marseille-P3236-mgs1413
s__Dasheen_mosaic_virus
Parabacteroides sp. Marseille-P3236-mgs0091
s__Lactobacillus_gasseri
s__Prevotella_buccalis
s__Dialister_micraerophilus
Bacteria-mgs0028
Number of vaginal deliveries
Coprococcus-mgs0796
Lachnospiraceae-mgs0463
Clostridiales bacterium KLE1615-mgs0481
s__Dialister_micraerophilus_t__Dialister_micraerophilus_unclassified
Percentile of forehead red area
Parabacteroides gordonii-mgs0020
Asparagine level
Cystathionine (μmol/L)
Vertical jump score
s__Streptococcus_anginosus
Ruthenibacterium lactatiformans-mgs0076
Arsenic (As) (μg/L)
s__Torque_teno_virus
Glutamine level
Anaerovoracaceae-mgs0594
Fusicatenibacter saccharivorans-mgs0468
s__Streptococcus_anginosus_t__Streptococcus_anginosus_unclassified
s__Torque_teno_virus_t__PRJNA70005
s__Pseudomonas_unclassified
Number of pregnancy
s__Escherichia_unclassified
Vagino-cervical microbiome
Metabolites
Physical fitness test
Facial skin feature
Lifestyle questionnaire
Female life history questionnaire
Fecal microbiome