RESEARCH ARTICLE

Vaginal microbiota and personal risk factors associated with HPV status conversion—A new approach to reduce the risk of cervical cancer?

Zhongzhou Yang[1☯*], Ye Zhang[2☯], Araceli Stubbe-Espejel[1☯], Yumei Zhao[1], Mengping Liu[3], Jianjun Li[2], Yanping Zhao[4], Guoqing Tong[5], Na Liu[1], Le Qi[1], Andrew Hutchins[6], Songqing Lin[2‡], Yantao Li[1‡]

1 BGI Genomics, BGI-Shenzhen, Shenzhen, China, **2** Department of Traditional Chinese Medicine, Women & Children Health Institute Futian Shenzhen, Shenzhen, China, **3** School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, China, **4** BGI-Shenzhen, Shenzhen Key Laboratory of Unknown Pathogen, Shenzhen, China, **5** Shouguang Hospital of Traditional Chinese Medicine, Reproduction Medicine Center Shanghai, China, **6** Department of Biology, Southern University of Science and Technology, Xueyuan Lu, Shenzhen, China

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* jacriyang5872@hotmail.com

Abstract

Vaginal microbiota (VMB) is associated with changes in Human papilloma virus (HPV) status, which consequently influences the risk of cervical cancer. This association was often confounded by personal risk factors. This pilot research aimed to explore the relationship between vaginal microbiota, personal risk factors and their interactions with HPV status conversion to identify the vaginal microbiota that was associated with HPV clearance under heterogeneous personal risk factors. A total of 38 women participated by self-collecting a cervicovaginal mucus (CVM) sample that was sent for metagenomics sequencing. Most of the participants also filled in personal risk factors questionnaire through an eHealth platform and authorized the use of their previous HPV genotyping results stored in this eHealth platform. Based on the two HPV results, the participants were grouped into three cohorts, namely HPV negative, HPV persistent infection, and HPV status conversion. The relative abundance of VMB and personal factors were compared among these three cohorts. A correlation investigation was performed between VMB and the significant personal factors to characterize a robustness of the panel for HPV status change using R programming. At baseline, 12 participants were HPV-negative, and 22 were HPV-positive. Within one year, 18 women remained HPV-positive, 12 were HPV-negative and 4 participants showed HPV clearance. The factors in the eHealth questionnaire were systematically evaluated which identified several factors significantly associated with persistent HPV infection, including age, salary, history of reproductive tract infection, and the total number of sexual partners. Concurrent vaginal microbiome samples suggest that a candidate biomarker panel consisting of *Lactobacillus gasseri*, *Streptococcus agalactiae*, and *Timona prevotella* bacteria, which may be associated with HPV clearance. This pilot study indicates a stable HPV

OPEN ACCESS

Citation: Yang Z, Zhang Y, Stubbe-Espejel A, Zhao Y, Liu M, Li J, et al. (2022) Vaginal microbiota and personal risk factors associated with HPV status conversion—A new approach to reduce the risk of cervical cancer? PLoS ONE 17(8): e0270521. https://doi.org/10.1371/journal.pone.0270521

Editor: Guangming Zhong, University of Texas Health Science Center at San Antonio, UNITED STATES

Received: December 14, 2021

Accepted: June 11, 2022

Published: August 9, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0270521

Data Availability Statement: Both metagenomics sequence data and personal data were also deposited in the CNGB Nucleotide Sequence

Archive (CNSA: https://db.cngb.org/cnsa; accession number CNP0002023).

Funding: This study was supported by funding from the Shenzhen Innovation Committee of Science and Technology (ZDSYS20200811144002008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

status-related vaginal microbe environment. To establish a robust biomarker panel for clinical use, larger cohorts will be recruited into follow-up studies.

Introduction

An abundance of species in the vaginal microbiota (VMB) has been associated with persistent infection with high-risk human papillomavirus (HPV) and the causative agent of cervical cancer [1] as well as personal factors [2]. VMB mainly includes the larger abundances of *Lactobacillus spp*. related to HPV negativity [3]. However, HPV infection was highly relevant to protective *Lactobacillus spp*. and pathogenic *Neisseria gonorrheae*, *Chlamydia trachomatis*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, *Streptococcus agalactiae* and *Timona prevotella* bacteria, which cause vaginosis [4]. Individual personal features belonging to precision medicine are beneficial to preventing persistent HPV infection or promoting HPV clearance [5].

We explored the associations between VMB and long-term HPV infection status (persistent infection or clearance) through metagenomic sequencing technology and consecutive HPV genotyping results through our digital eHealth platform. The eHealth platform was also used to collect various types of individual factors for reducing heterogeneity. Using this digital eHealth platform, our team systematically collected the personal factors that might be associated with HPV infection from the literature [6]. These factors included five categories: demographics (e.g., age), personal disease history [7], lifestyle behavior on malnutrition [8], sexual history [9–11] on the number of sexual partners and substance abuse on smoking habits [12].

To the best of our knowledge, this study was the first to identify VMB biomarkers by performing a systematic exploration of the potential confounding variables of HPV infection [2]. After obtaining metagenomics sequences and other factors through the eHealth platform, a correlation approach was utilized to explore the association between the candidate biomarkers and personal factors. We define stable microbiomes as biomarkers that are not influenced by the status of other crucial factors. The correlation p-value is utilized to select the stable biomarker panels as overlapping for each category.

Materials and methods

Participants recruitment and sample collection

This research was approved by the ethics committee of the Institutional Review Board at Beijing Genomics Institution (BGI-IRB 21054). This research is recorded with www.chictr.org.cn, ChiCTR2100049221. The recruitment of participants for this study began on May 25th, 2021 and was carried out in a community setting in Shenzhen, Mainland China. Eligible participants were nonpregnant, nonlactating women who had sex at least once in their lifetime. Permission was given by the participants for the research team to use their eHealth data for both health record data including current and previous (within the last 12 months) HPV test results, as well as social personal factors. Based on the HPV test results, participants were grouped into three cohorts: HPV-negative (both samples were HPV negative), HPV-negative conversion (conversion from HPV positive to HPV negative), and HPV-positive subjects (both samples were HPV positive, suggesting persistent infection).

Once the subjects filled in their information in a registration form (S1 Text), they received a metagenomic self-sampling kit via mail, including clear instructions. Participants were requested to abstain from vaginal intercourse 24 hours before sampling, to wait for at least three days after menstrual blood was cleared and to avoid using vaginal douches and any vaginally administered medical treatments [13–16]. A sample of vaginal mucus was collected by inserting a swab into the vagina. The swab was then stirred/placed inside a special tube with a DNA preservative solution N-octylpyridinium bromide (NOPB) [17,18]. The participants were then instructed to close the tube and place it in a plastic bag until the pick-up was arranged, and the sample was collected at room temperature for up to at least 14 days. Both the tube and the bag had a barcode/QR code for identification.

Personal factor and eHealth platform

PROs (personal record outcomes) are defined as reports directly from the participants about the health condition status of the patient's response without interpretation or amendment by doctor or anyone else. Participants in the study were required to upload their personal PROs, which were divided into five categories with 32 personal factors (S2 Text), on the eHealth platform (CanSeq). It covers several factors, including demographics, medical history, lifestyle (S3–S6 Text), sexual history and behavior and substance abuse factors. The eHealth platform enables noninteractive support for the participants for multiple purposes. First, video and written instructions were provided on the eHealth platform to guide the participants for sample collection. Second, HPV infection records since 2016 are recorded, as authorized by users. Third, the registration of the participant´s information, including evaluated eligibility for participating in the screening program and for the collection of PROs for analytical purposes.

Laboratory tests

After the metagenomic self-sampling kits were returned, they were sent to the China National GeneBank DataBase (CNGB). DNA was extracted for 38 samples as formerly mentioned [19–21]. Furthermore, DNA libraries were prepared as one paired-end (PE) with 350 bp insert size for individual sample [19]. A length of each read is from 75bp at stage I to 90 bp at stage II. A shotgun of metagenomic was sequenced on BGISEQ-500 platform that is equivalent with other sequencing platforms [22–24]. Data (S7 Text) analysis was carried out using an onsite pipeline, and profiles were uploaded on the online cloud pipeline [25].

The HPV test results of the participants were obtained from the eHealth platform where the SeqHPV test (BGI Shenzhen, Shenzhen, China) results of the participants were stored. The SeqHPV test is a kit to detect HPV infection in female cervical exfoliated epithelial cells by using a combinatorial probe-anchor synthesis (cPAS) sequencing approach [26]. It is utilized to detect 2 low-risk types of HPV (types: 6, 11) and 14 high-risk HPV types (types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) [27].

Statistical analysis

After the samples were received and sequenced, the relative abundance of the 16 VMBs was compared using *Lactobacillus* as a reference in each cohort. To differentiate the VMB profile, the relative level of microorganism abundance was applied for each cohort. The significant personal factors and microbiome were expressed as a number for categorical variables and mean ± SD for continuous variables. Analysis of variance (ANOVA) was used to compare the demographic factors. A p-value < 0.05 was considered statistically significant.

After conducting abundance and personal factor analysis, Pearson correlation analysis was applied to link the key personal indicators (S1 Table) and candidate microorganism biomarkers by linear regression for these three cohorts [28]. Student's t-test was used to determine the significance of correlation for the microorganisms in the VMB versus the HPV infection. The resulting p-values for each microorganism were used to select candidate biomarkers for all cross-comparisons. A p-value < 0.05 was considered statistically significant for each test.

Results

Participant recruitment, and HPV cohorts grouping

Forty-four participants were invited to join the pilot study. Initially, 38 joined this study, but four dropped because they declined to provide their personal data (Fig 1). Thus, a total of 34 participants qualified for this study. The 34 participants completed the vaginal mucus samples and provided HPV status and personal data (S2 Table). Based on the HPV infection records within the last 12 months, the participants were grouped into three cohorts: 12 participants

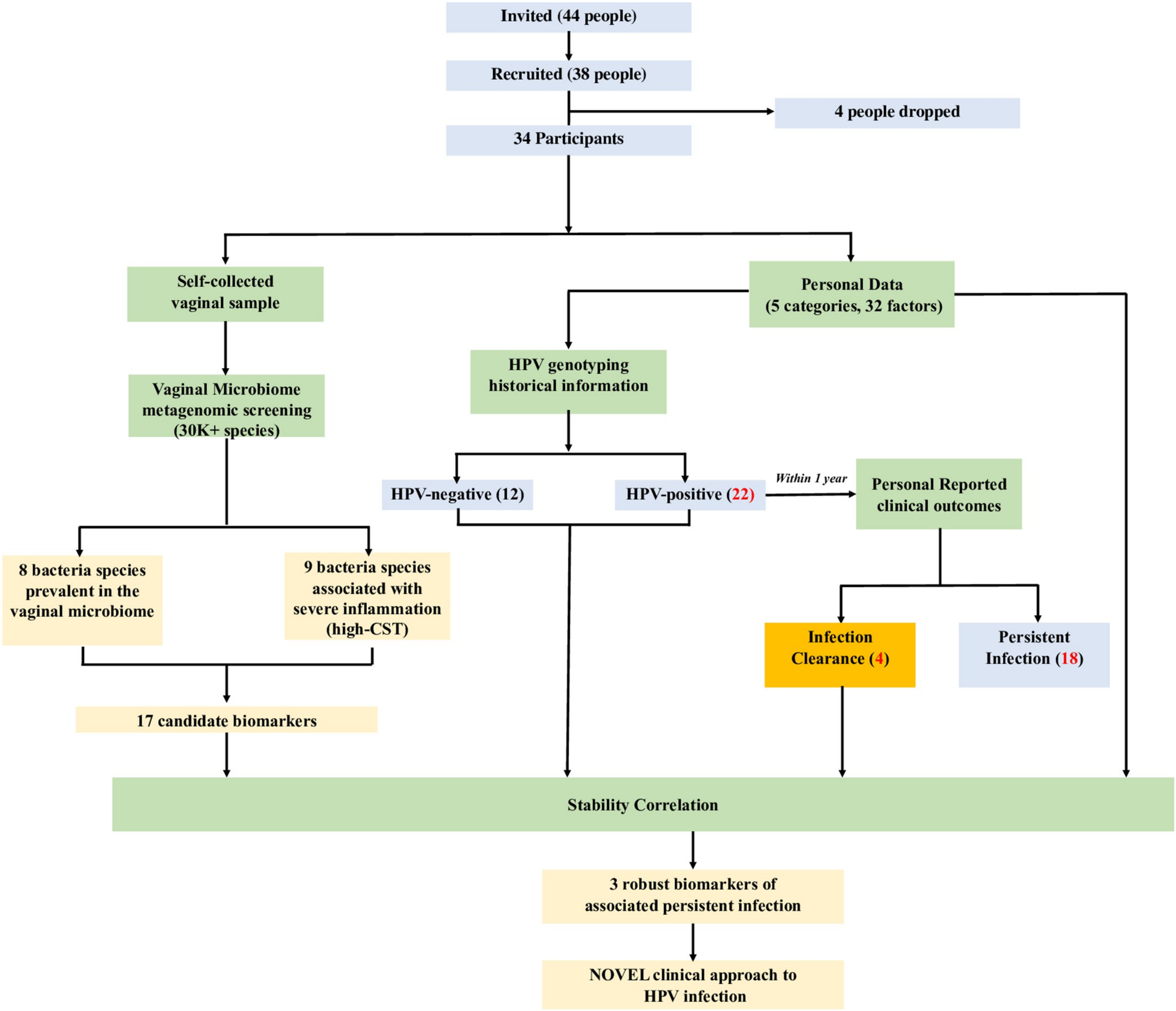

Invited (44 people)
Recruited (38 people)
4 people dropped
34 Participants
Self-collected vaginal sample
Personal Data (5 categories, 32 factors)
Vaginal Microbiome metagenomic screening (30K+ species)
HPV genotyping historical information
HPV-negative (12)
HPV-positive (22)
Within 1 year
Personal Reported clinical outcomes
8 bacteria species prevalent in the vaginal microbiome
9 bacteria species associated with severe inflammation (high-CST)
Infection Clearance (4)
Persistent Infection (18)
17 candidate biomarkers
Stability Correlation
3 robust biomarkers of associated persistent infection
NOVEL clinical approach to HPV infection

Fig 1. Flowchart for identifying biomarkers between vaginal microbiota and HPV status. 30K = 30,000; HPV = Human papillomavirus; CST = community state type.

https://doi.org/10.1371/journal.pone.0270521.g001

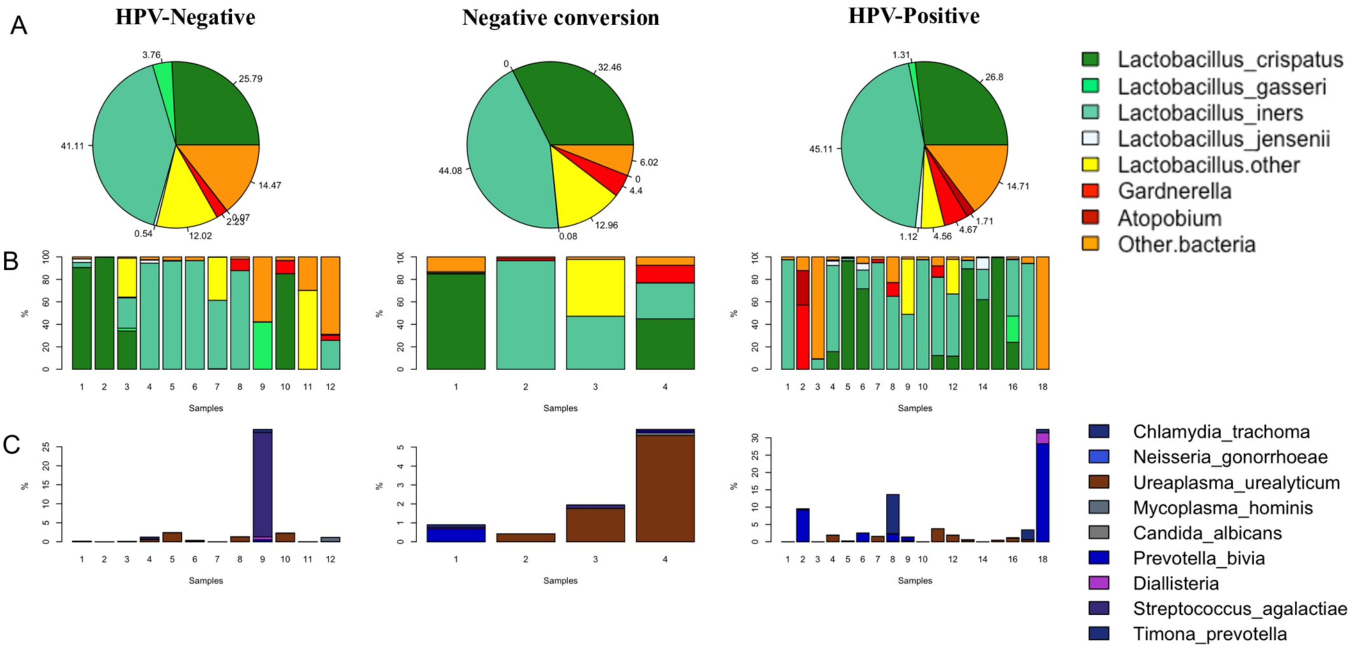


Fig 2. Relative abundance (%) of species of microbes in the three cohorts. Legend: **(A)** Pie charts show the relative microorganism abundance between the three cohorts. Proportion was calculated from the average value of abundance for each group by CST type. **(B)** Bar charts showing the proportion of dominant species in each sample. Selected microorganism level was selected from the CST type to show the relative abundance and characterization. **(C)** Bar charts showing the proportion of pathogenic microorganism species as indicated in the key.

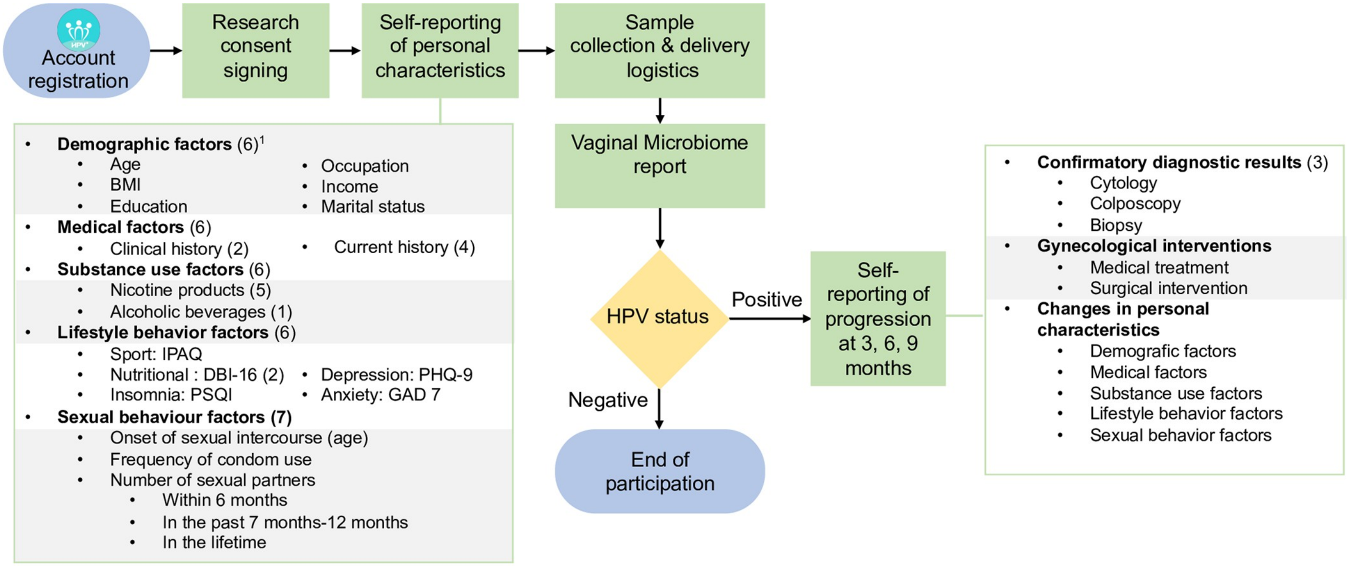
https://doi.org/10.1371/journal.pone.0270521.g002

were placed into the HPV-negative cohort (i.e., no HPV infection within the last year), 18 in the persistent HPV-positive cohort (persistent infection suggested by two HPV-positive results spaced 12 months apart) and 4 in the HPV positive-to-negative conversion cohort (i.e., Previously positive, but the most recent test was negative), and there were no new HPV-positive categories (i.e., Previously negative, but the most recent test was positive). Finally, we obtained 17 biomarkers to explore the relative abundance and stability correlation with personal data for five category groups. Both metagenomics sequence data and personal data were also deposited in the CNGB Nucleotide Sequence Archive (CNSA: https://db.cngb.org/cnsa; accession number CNP0002023).

Vaginal microbiome

In the pilot stage, we focused on 17 VMBs (vaginal microbiomes), including nine community state types (CSTs) and eight gynecological diseases from the literature [29–31] through metagenomic sequencing. Overall, metagenomic sequencing identified 17 species in 9 clades (Fig 2A and 2B and S2 Text). Lactobacillus genus microorganisms were predominant in the VMB of the three cohorts, composing over 80% in most samples, agreeing closely with the patterns in previous vaginal samples [32]. In particular, *Lactobacillus iners* was identified as the predominant species among all three cohorts in this study (Fig 2A & 2B), with *Lactobacillus crispatus* being the second most abundant.

Pathogenic *Gardnerella spp* had a higher presence in HPV current or past infections. However, *atopobium* was only substantially observed in HPV-positive samples. On the other hand, all of the nine previously were identified as pathogenic gynecological infection-causing pathogens on *Trachoma chlamydia*, *Neisseria gonorrheae*, *Microureaplasma*, *Mycoplasma hominis*, *Candida albicans*, *Prevotella bivia*, *Diallisteria*, *Streptococcus agalactiae* and *Timona prevotella*.



¹The number of questions is shown in the parenthesis

Fig 3. Translational eHealth platform flowchart for the collection for participant-reported outcomes (PROs). Legend: IPAQ = international physical activity questionnaires. DBI = diet balance index. PSQI = Pittsburgh sleep quality index. PHQ-9 = Patient Depression Questionnaire-9. GAD 7 = Generalized Anxiety Disorder 7.

https://doi.org/10.1371/journal.pone.0270521.g003

They were presented in only minor proportions at 0.40% ± 2.45% among the three cohorts (Fig 2C).

Translational eHealth platform

The eHealth application is used to interact with the participants to manage HPV test results and to collect three personal character groups and two participant-reported outcome (PRO) groups related to HPV infections (Fig 3). Participants were requested to answer a list of questions related to several factors, including simple biometrics (age, body mass index (BMI), demographic state (education, occupation, salary and marital status), medical history (six factors), substance abuse (six factors), lifestyle (six factors) and sexual history and behavior (six factors), which may affect the risk of being infected with HPV or other pathogenic microorganisms (Fig 3).

When the participants were positive for any of the HPV serotypes covered by the HPV test, they were prompted to update the eHealth questionnaires every three months. Then, the program prompts the participant to provide updates on PROs, seroconversion period, additional confirmatory diagnostic test results, and updates on their medical history, immunization (HPV vaccination) history, lifestyle changes such as starting new sports activities, changes to their usual diet, quality of sleep and psychological status, substance abuse smoking, alcohol, sexual history and behavior.

Personal risk factors: Precision medicine

The results of the 32 PROs and their statistical association with the three types of HPV status are shown in Fig 4 (S3 Table). Demographic factors were significantly correlated with HPV status on age and salary, while education, career, BMI and matrial status were weakly correlated (Fig 4A). Fig 4B shows the medical history, including history of disease or current infection. Both reproductive tract infection (RTI) and a history of consanguineous hereditary or nonhereditary cancer seem potentially related to HPV-positive cases. Fig 4C shows the

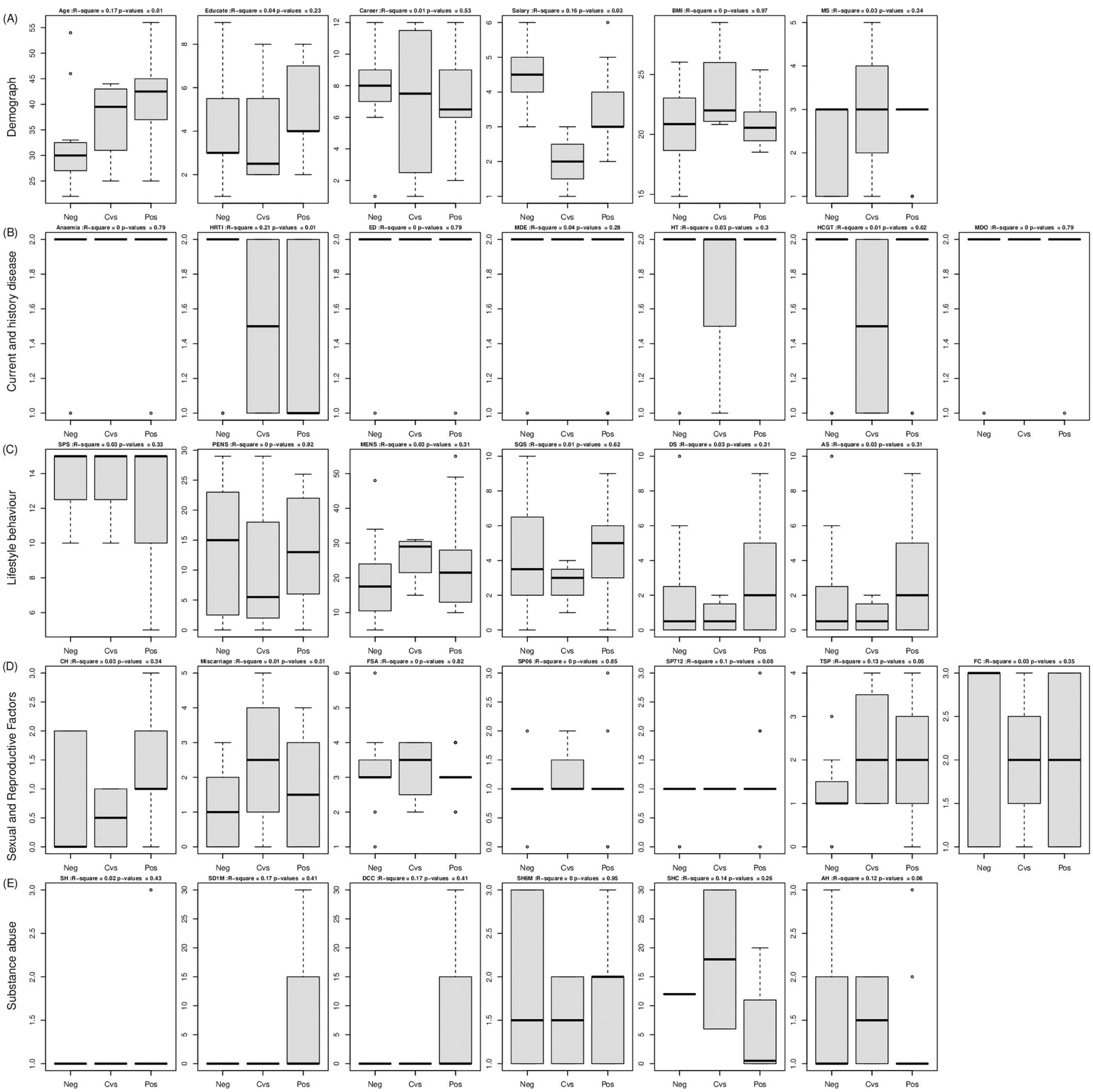


Fig 4. Personal factors from the PROs of the participants and relation to HPV-negative, negative conversion, and HPV-positive factors. Legend: 32 personal factors from 5 categories on three types of status. **(A)** Six demographical factors including age, educate, career and etc; MS = Marital status. **(B)** Seven medical history factors including history of disease or current infection; HRTI = History of Reproductive Tract Infection, ED = Endocrine disease, MDE = Metabolic disease, HT = History of tumor, HCGT = History of consanguineous tumor, MDO = Mental disorder. **(C)** Six behavior factors and their association with the HPV status; SPS = Sport scores, PENS = PE Nutrient scores (well nourished), MENS = ME Nutrient scores (malnourishment), SQS = Sleep quality scores, DS = Depression scores, AS = Anxiety scores. **(D)** Seven sexual and reproductive factors; CH = Childbearing history, FSA = First sexual age, SP06 = Sexual partners 0-6M, SP712 = Sexual partners 7-12M, TSP = Total sexual partners, FC = Frequency of condom. **(E)** Six substance abuse factors. SH = Smoking habit, SD1M = Smoking days within 1M, DCC = Daily cigarette consumption, SH6M = Second hand more than 6M, SHC = Secondhand cigarette, AH = Alcohol habit.

https://doi.org/10.1371/journal.pone.0270521.g004

behavior factors and their association with HPV status. Malnourishment was a pseudosignificant factor for HPV infection. However, the strongest correlation between HPV status and the total number of sexual partners was also correlated (Fig 4D). However, the number of days smoking and daily cigarette consumption were not correlated with HPV status. This suggests that smoking and alcohol consumption may ultimately be indirect demographic factors (Fig 4E).

Significant personal factors and bacteria

The baseline personal significant factors and metagenomic data are shown as eleven items in **Table 1**. Among the HPV-negative cohort, negative conversion and HPV-positive subjects, both age and history of reproductive tract infection had a consistent pattern. Another two significant demographic and behavioral factors were the salary range and the total number of sex partners with an inconsistent pattern. The negative HPV test results tended to be associated with higher salaries. Cohorts of participants within the lowest salary range and with the largest number of total sex partners showed greater seroconversion. HPV-negative subjects had the lowest number of total sex partners.

After adjusting for age, salary, history of reproductive tract infection and the total number of sexual partners, the metagenomics data showed that both *Lactobacillus jensenii* and *Streptococcus agalactiae* were a relatively abundant part of the VMB, and another 5 types were pseudosignificant due to the limited sample size in this pilot study. *Lactobacillus jensenii*, for example, had a relatively higher proportion in the HPV-positive group and a reduced proportion in the seroconversion group. The presence of *Streptococcus agalactiae* seemed to have a correlation between HPV-negative and HPV-positive seroconversion.

Correlation between personal factors and microbiome

To determine the stable potential candidate biomarkers, a correlation analysis was conducted between four significant personal factors and seven microorganism species, as shown in Fig 5.

Table 1. Significant or pseudo-significant characteristics of the participants.

Risk factors	HPV-Negative (12)	Negative conversion (4)	HPV-Positive (18)	p-Value [a]
Personal				
Age (year)	31.9 ± 9.3	37.0 ± 8.5	40.4 ± 8.8	0.01
Salary[b] (¥)	4.6 ± 1.0	2.0 ± 1.0	3.4 ± 1.1	0.03
History of reproductive tract infection[c]	1.8 ± 0.4	1.5 ± 0.6	1.3 ± 0.5	0.01
Total sexual partners[d]	1.2 ± 0.9	2.3 ± 1.5	2.1 ± 1.0	0.049
Microorganism type				
Lactobacillus gasseri	3.8 ± 12.1	0 ± 0	1.3± 5.5	0.06
Lactobacillus jensenii	0.5 ± 1.1	0.1± 0.2	1.1 ± 2.8	<0.01
Atopobium vaginae	0.1± 0.2	0 ± 0	1.7 ± 7.2	0.06
Mycoplasma hominis	0.1± 0.3	0 ± 0	0 ± 0	0.09
Prevotella bivia	0.1± 0.2	0.2 ± 0.3	2.4± 6.8	0.07
Streptococcus agalactiae	2.3± 7.9	0.1± 0.1	0 ± 0	<0.05
Timona_prevotella	0.2 ± 0.3	0 ± 0.1	0.9 ± 2.7	0.06

Legend: Values are mean ± SD.

[a] Statistical difference by ANOVA (Analysis of Variance).

[b] 1: < 1000 CND; 2:1000~3000 CND; 3: 3000~5000 CND; 4: 5000~10000 CND.

[c] 1 = Yes, 2 = No, the higher the values, the lesser the probability.

[d] 3 = 3–5 partners.

[e] Adjusted for personal variables. Since this is a limited sample size, statistical difference was computed by comparing negative and positive cohorts.

https://doi.org/10.1371/journal.pone.0270521.t001

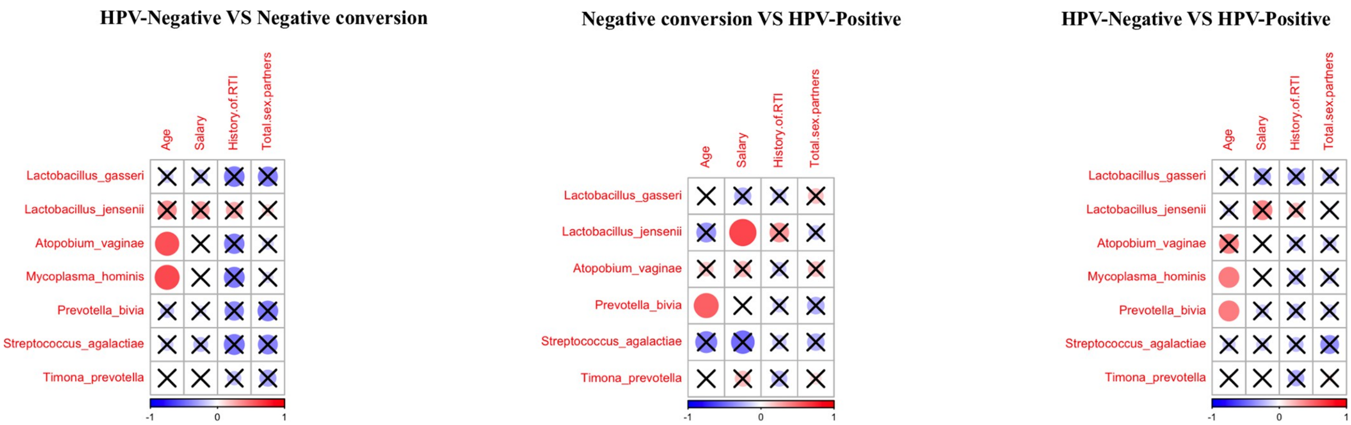


Fig 5. Association between personal indicators the candidate biomarkers. Legend: Correlation coefficients between four potential biomarkers and personal indicators in HPV-negative vs negative-conversion, negative-conversion vs HPV-positive, and HPV-negative vs HPV-positive cohorts. Red and blue represent positive and negative associations. Crosses represent no significant correlation (p-value > 0.05). The size of the circle represents the R-value of the personal factors and the microorganisms calculated from the linear regression.

https://doi.org/10.1371/journal.pone.0270521.g005

Age has a significant association with *atopobium vaginae* and *mycoplasma hominis* in HPV-negative samples; *atopobium vaginae* and *prevotella bivia* are present in the seroconversion cases; *mycoplasma hominis* and *prevotella bivia* are abundant in the HPV-positive group. *Mycoplasma hominis* was not found in the seroconverted cohort. Other associations between personal factors and vaginal bacteria were not significant.

To identify robust biomarkers of correlation with gynecological health, finding a stable biomarker is fundamental. To increase the potential of using microbiome analysis as a useful tool in the community setting, the overlap was theoretically defined of the microorganism present within all three cohorts. *Lactobacillus gasseri*, *Streptococcus agalactiae*, and *Timona prevotella* were identified as candidate biomarkers of cervicovaginal health and differentiate HPV status.

Discussion

This study explores the effect of the presence of different microorganisms in the vaginal microbiome of HPV-negative, HPV-positive to HPV-negative individuals and persistent HPV-positive individuals. In regard to the microorganisms that were found in vaginal mucus samples, the presence of species from the *Lactobacillus* genus dominated the microbiome, with notable representation of the *Lactobacillus iners* species. Notably, the three bacteria *Lactobacillus gasseri*, *Streptococcus agalactiae*, and *Timona prevotella* were differentially correlated to the three cohorts analyzed in this study. Overall, five microorganisms are beneficial to humans, including *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus iners*, and *Lactobacillus jensenii*; 12 are pathogenic, including *Gardnerella vaginalis*, *Atopobium vaginae*, *Trachoma chlamydia*, *Neisseria gonorrheae*, *Microureaplasma*, *Mycoplasma hominis*, *Candida albicans*, *Prevotella bivia*, *Diallisteria*, *Streptococcus agalactiae* and *Timona prevotella*, among these 17 microorganisms. Overall, we did not find that an increased level of pathogenic bacteria was correlated with HPV status, but changes in the balance of the normal vaginal microbiome were associated with HPV infection.

Our study agrees with previous studies showing that Lactobacillus spp. are highly abundant in the vaginal microbiome [33–35]. However, the proportion of anerobic bacteria was quite discrepant with the lower abundance. For example, *Ureaplasma urealyticum* was low at 0.55% in the HPV-negative cohort, 1.95% in the negative conversion cohort and 0.67% in the HPV-positive cohort.

We also took into consideration the 32 factors that belong to five categories, namely, demographic, medical history, lifestyle, sexual history and behavior, and substance abuse factors. Among these factors, four of them were statistically significant as age, salary, history of reproductive tract infection and total sexual partners. Specifically, history of reproductive tract infection was accounted for the association to identify biomarker panels. Other plausible factors have not accounted for the association likely, number of kids, age that women have babies, mode of delivery and HIV infection because of insignificance.

Considering the personal factors from the eHealth platform, this study first identified these three biomarkers via a correlation study. The eHealth platform is a user-friendly mobile application program that is more cost-efficient than any other kind of management and requires personal attention to every participating individual in a health-related screening program. For convenience, there is the acceptance of the Privacy Policy and research's Informed Consent. Moreover, it became the platform to report the results of the participant's clinical test, and technical assistance was provided when required. The personal feature dataset can also be used to invite the participants to enroll in additional related observational or interventional studies, such as VMB-probiotic treatment studies, VMB-screening feasibility trials and longitudinal multicenter VMB invasion research.

The burden of cervical cancer can be effectively reduced to take measures on the diagnosis, probiotic, corresponding behavior intervention and assess feasibility for future research directions. First, these three biomarkers can consist of an optimized diagnosis panel for VMB. Then, a biomarker panel can potentially be developed into probiotic bacteria for treatment. Third, the eHealth platform was potentially for lifestyle intervention, particularly for significant personal factors with the aim of minimizing the need for intervention through easy-to-follow simple instructions. One feature of a dynamic eHealth platform is that personalized feedback, a free one-to-one online medical consultation, and education training can be provided for participant-centered care. Finally, a biomarker will be evaluated in multiple centers to develop a product to reduce the risk of cervical cancer.

Strengths and weaknesses

In this study, we have three strengths. An eHealth platform is able to gather a personal dataset to explore the stability of biomarkers. In addition, the existing metagenomic tools provide an opportunity to carry out comprehensive analyses and identify even slight variations in the abundance of microorganisms in the VMB. While 16S rRNA sequencing is traditionally used to identify the microorganism composition of the VMB, the approach is not suitable for identifying an ampler diversity of biological entities and their interrelations. Metagenomics, on the other hand, is a powerful tool that has been used to carry out broader genus searches as well as biomarker identification for drug development, as it provides one of the most compatible techniques to detect microorganisms with high reproducibility and robust reliability.

Self-sampling for vaginal mucus and vaginal epithelium cells is a relatively new feature of gynecological screening processes, which may help to remove some of the personal barriers that limit the participation of women in cervical health screening programs and expand the reach of public health interventions to remote regions by eliminating the need to have the sampling performed by a specialist in a clinical setting. Studies have shown that self-collected samples yield results that are comparable to those collected by healthcare professionals [36].

There is one major concern in this study. The sample size of this pilot study limits the statistical power of the results. The sample might not be demographically representative. To reduce the heterogeneity, an eHealth platform was utilized to deep systematically collect personal factors, which were further utilized to identify the robust biomarker panels. Additionally, our future formal study will be the larger number of subjects to elucidate mechanisms of VMB biomarker panel.

Conclusions

This work aimed to identify novel microorganism-based biomarkers of HPV infection, discerning among its different stages. *Metagenomic studies have shown that Lactobacillus gasseri*, *Streptococcus agalactiae and Timona prevotella* bacteria and their relative abundances are markedly different among different cohorts of HPV infection. An enhanced vaginal microbiota biomarker panel could be created one potential robust clinical tool for HPV acquisition, persistence or clearance, since their identification procedure is one of the biggest challenges in the ongoing colposcopy.

Supporting information

S1 Table. Clinical indicator.
(XLSX)

S2 Table. Clinical data completion.
(XLSX)

S3 Table. Participants clinical factor output).
(XLSX)

S1 Text. Recruitment procedure.
(DOCX)

S2 Text. Clinical factors by questionnaire.
(DOCX)

S3 Text. Personal factors on sport (IPAQ).
(PDF)

S4 Text. Personal factors on insomnia (PSQI).
(PDF)

S5 Text. Personal factors on Depression (PHQ-9).
(PDF)

S6 Text. Personal factors on Anxiety (GAD-7).
(PDF)

S7 Text. Metagenomics.
(DOCX)

Acknowledgments

We acknowledge the BGI IT Team (Chungen Zhang, Hao Zhou and Weiwei Bai) for developing the eHealth platform and Kaiye Cai for her bioinformatics assistance.

Author Contributions

Conceptualization: Zhongzhou Yang, Ye Zhang, Yantao Li.

Data curation: Araceli Stubbe-Espejel, Yumei Zhao, Jianjun Li, Yantao Li.

Formal analysis: Zhongzhou Yang, Mengping Liu, Yanping Zhao, Andrew Hutchins.

Funding acquisition: Yantao Li.

Investigation: Zhongzhou Yang, Araceli Stubbe-Espejel, Yanping Zhao, Guoqing Tong.

Methodology: Zhongzhou Yang, Araceli Stubbe-Espejel, Mengping Liu.

Resources: Yumei Zhao, Jianjun Li, Na Liu, Le Qi, Yantao Li.

Software: Zhongzhou Yang, Le Qi, Songqing Lin.

Visualization: Zhongzhou Yang.

Writing – original draft: Zhongzhou Yang.

Writing – review & editing: Zhongzhou Yang, Araceli Stubbe-Espejel, Yanping Zhao, Andrew Hutchins.

References

1. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019; 393(10167):169–82. Epub 2019/01/15. https://doi.org/10.1016/S0140-6736(18)32470-X PMID: 30638582.
2. Mitra A, MacIntyre DA, Marchesi JR, Lee YS, Bennett PR, Kyrgiou M. The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? Microbiome. 2016; 4(1):58. Epub 2016/11/03. https://doi.org/10.1186/s40168-016-0203-0 PMID: 27802830; PubMed Central PMCID: PMC5088670.
3. Laniewski P, Ilhan ZE, Herbst-Kralovetz MM. The microbiome and gynaecological cancer development, prevention and therapy. Nat Rev Urol. 2020; 17(4):232–50. Epub 2020/02/20. https://doi.org/10.1038/s41585-020-0286-z PMID: 32071434.
4. Lev-Sagie A, Goldman-Wohl D, Cohen Y, Dori-Bachash M, Leshem A, Mor U, et al. Vaginal microbiome transplantation in women with intractable bacterial vaginosis. Nat Med. 2019; 25(10):1500–4. Epub 2019/10/09. https://doi.org/10.1038/s41591-019-0600-6 PMID: 31591599.
5. Yates LR, Seoane J, Le Tourneau C, Siu LL, Marais R, Michiels S, et al. The European Society for Medical Oncology (ESMO) Precision Medicine Glossary. Ann Oncol. 2018; 29(1):30–5. Epub 2017/11/16. https://doi.org/10.1093/annonc/mdx707 PMID: 29140430.
6. Winer RL, Hughes JP, Feng Q, O'Reilly S, Kiviat NB, Holmes KK, et al. Condom use and the risk of genital human papillomavirus infection in young women. N Engl J Med. 2006; 354(25):2645–54. Epub 2006/06/23. https://doi.org/10.1056/NEJMoa053284 PMID: 16790697.
7. Rachana KC, Giri R. Knowledge regarding cervical cancer among undergraduate female students at a selected college of Lalitpur, Nepal. Can Oncol Nurs J. 2019; 29(3):184–8. Epub 2020/01/23. https://doi.org/10.5737/23688076293184188 PMID: 31966010; PubMed Central PMCID: PMC6970460.
8. Harrowfield J, Isenring E, Kiss N, Laing E, Lipson-Smith R, Britton B. The Impact of Human Papillomavirus (HPV) Associated Oropharyngeal Squamous Cell Carcinoma (OPSCC) on Nutritional Outcomes. Nutrients. 2021; 13(2). Epub 2021/02/10. https://doi.org/10.3390/nu13020514 PMID: 33557340; PubMed Central PMCID: PMC7916068.
9. Vinodhini K, Shanmughapriya S, Das BC, Natarajaseenivasan K. Prevalence and risk factors of HPV infection among women from various provinces of the world. Arch Gynecol Obstet. 2012; 285(3):771–7. Epub 2011/12/14. https://doi.org/10.1007/s00404-011-2155-8 PMID: 22159694.
10. Zhao FH, Tiggelaar SM, Hu SY, Xu LN, Hong Y, Niyazi M, et al. A multi-center survey of age of sexual debut and sexual behavior in Chinese women: suggestions for optimal age of human papillomavirus vaccination in China. Cancer Epidemiol. 2012; 36(4):384–90. Epub 2012/03/02. https://doi.org/10.1016/j.canep.2012.01.009 PMID: 22377277; PubMed Central PMCID: PMC5523958.
11. Chen J, Gopala K, Akarsh PK, Struyf F, Rosillon D. Prevalence and Incidence of Human Papillomavirus (HPV) Infection Before and After Pregnancy: Pooled Analysis of the Control Arms of Efficacy Trials of HPV-16/18 AS04-Adjuvanted Vaccine. Open Forum Infect Dis. 2019; 6(12):ofz486. Epub 2019/12/12. https://doi.org/10.1093/ofid/ofz486 PMID: 31824976; PubMed Central PMCID: PMC6892569.
12. Brotman RM, He X, Gajer P, Fadrosh D, Sharma E, Mongodin EF, et al. Association between cigarette smoking and the vaginal microbiota: a pilot study. BMC Infect Dis. 2014; 14:471. Epub 2014/08/30. https://doi.org/10.1186/1471-2334-14-471 PMID: 25169082; PubMed Central PMCID: PMC4161850.
13. Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UM, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. Sci Transl Med. 2012; 4(132):132ra52. Epub 2012/05/04. https://doi.org/10.1126/scitranslmed.3003605 PMID: 22553250; PubMed Central PMCID: PMC3722878.

14. Belinson JL, Wang G, Qu X, Du H, Shen J, Xu J, et al. The development and evaluation of a community based model for cervical cancer screening based on self-sampling. Gynecol Oncol. 2014; 132(3):636–42. Epub 2014/01/21. https://doi.org/10.1016/j.ygyno.2014.01.006 PMID: 24440471.

15. Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. Microbiome. 2014; 2:23. Epub 2014/07/24. https://doi.org/10.1186/2049-2618-2-23 PMID: 25053998; PubMed Central PMCID: PMC4106219.

16. Song SD, Acharya KD, Zhu JE, Deveney CM, Walther-Antonio MRS, Tetel MJ, et al. Daily Vaginal Microbiota Fluctuations Associated with Natural Hormonal Cycle, Contraceptives, Diet, and Exercise. mSphere. 2020; 5(4). Epub 2020/07/10. https://doi.org/10.1128/mSphere.00593-20 PMID: 32641429; PubMed Central PMCID: PMC7343982.

17. Han M, 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(1):43. Epub 2018/02/28. https://doi.org/10.1186/s40168-018-0429-0 PMID: 29482661; PubMed Central PMCID: PMC5828344.

18. Qian XB, Chen T, Xu YP, Chen L, Sun FX, Lu MP, et al. A guide to human microbiome research: study design, sample collection, and bioinformatics analysis. Chin Med J (Engl). 2020; 133(15):1844–55. Epub 2020/07/01. https://doi.org/10.1097/CM9.0000000000000871 PMID: 32604176; PubMed Central PMCID: PMC7469990.

19. 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(8):834–41. Epub 2014/07/07. https://doi.org/10.1038/nbt.2942 PMID: 24997786.

20. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012; 490(7418):55–60. Epub 2012/10/02. https://doi.org/10.1038/nature11450 PMID: 23023125.

21. Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. Appl Environ Microbiol. 1997; 63(7):2802–13. Epub 1997/07/01. https://doi.org/10.1128/aem.63.7.2802-2813.1997 PMID: 9212428; PubMed Central PMCID: PMC168577.

22. 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(3):1–8. Epub 2018/01/03. https://doi.org/10.1093/gigascience/gix133 PMID: 29293960; PubMed Central PMCID: PMC5848809.

23. 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(10). Epub 2018/09/08. https://doi.org/10.1093/gigascience/giy107 PMID: 30192933; PubMed Central PMCID: PMC6177736.

24. 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(5). Epub 2018/05/16. https://doi.org/10.1093/gigascience/giy055 PMID: 29762673; PubMed Central PMCID: PMC5967468.

25. Jie Z, Chen C, Hao L, Li F, Song L, Zhang X, et al. Life History Recorded in the Vagino-cervical Microbiome Along with Multi-omics. Genomics Proteomics Bioinformatics. 2021. Epub 2021/06/13. https://doi.org/10.1016/j.gpb.2021.01.005 PMID: 34118463.

26. Fehlmann T, Reinheimer S, Geng C, Su X, Drmanac S, Alexeev A, et al. cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. Clin Epigenetics. 2016; 8:123. Epub 2016/11/30. https://doi.org/10.1186/s13148-016-0287-1 PMID: 27895807; PubMed Central PMCID: PMC5117531.

27. Du H, Duan X, Liu Y, Shi B, Zhang W, Wang C, et al. An evaluation of solid versus liquid transport media for high-risk HPV detection and cervical cancer screening on self-collected specimens. Infect Agent Cancer. 2020; 15(1):72. Epub 2020/12/10. https://doi.org/10.1186/s13027-020-00333-4 PMID: 33292341; PubMed Central PMCID: PMC7706049.

28. Li J, Li J, Wang H, Qi LW, Zhu Y, Lai M. Tyrosine and Glutamine-Leucine Are Metabolic Markers of Early-Stage Colorectal Cancers. Gastroenterology. 2019; 157(1):257–9 e5. Epub 2019/03/20. https://doi.org/10.1053/j.gastro.2019.03.020 PMID: 30885779.

29. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. J Infect Dis. 2014; 210(11):1723–33. Epub 2014/06/20. https://doi.org/10.1093/infdis/jiu330 PMID: 24943724; PubMed Central PMCID: PMC4296189.

30. 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(1):29–37. Epub 2017/01/15. https://doi.org/10.1016/j.immuni.2016.12.013 PMID: 28087240; PubMed Central PMCID: PMC5270628.

31. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011; 108 Suppl 1:4680–7. Epub 2010/06/11. https://doi.org/10.1073/pnas.1002611107 PMID: 20534435; PubMed Central PMCID: PMC3063603.

32. Laniewski P, Barnes D, Goulder A, Cui H, Roe DJ, Chase DM, et al. Linking cervicovaginal immune signatures, HPV and microbiota composition in cervical carcinogenesis in non-Hispanic and Hispanic women. Sci Rep. 2018; 8(1):7593. Epub 2018/05/17. https://doi.org/10.1038/s41598-018-25879-7 PMID: 29765068; PubMed Central PMCID: PMC5954126.

33. Ilhan ZE, Laniewski P, Thomas N, Roe DJ, Chase DM, Herbst-Kralovetz MM. Deciphering the complex interplay between microbiota, HPV, inflammation and cancer through cervicovaginal metabolic profiling. EBioMedicine. 2019; 44:675–90. Epub 2019/04/28. https://doi.org/10.1016/j.ebiom.2019.04.028 PMID: 31027917; PubMed Central PMCID: PMC6604110.

34. Cheng L, Norenhag J, Hu YOO, Brusselaers N, Fransson E, Ahrlund-Richter A, et al. Vaginal microbiota and human papillomavirus infection among young Swedish women. NPJ Biofilms Microbiomes. 2020; 6 (1):39. Epub 2020/10/14. https://doi.org/10.1038/s41522-020-00146-8 PMID: 33046723; PubMed Central PMCID: PMC7552401.

35. Mitra A, MacIntyre DA, Ntritsos G, Smith A, Tsilidis KK, Marchesi JR, et al. The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. Nat Commun. 2020; 11(1):1999. Epub 2020/04/26. https://doi.org/10.1038/s41467-020-15856-y PMID: 32332850; PubMed Central PMCID: PMC7181700.

36. Arbyn M, Smith SB, Temin S, Sultana F, Castle P, Collaboration on S-S, et al. Detecting cervical precancer and reaching underscreened women by using HPV testing on self samples: updated meta-analyses. BMJ. 2018; 363:k4823. Epub 2018/12/07. https://doi.org/10.1136/bmj.k4823 PMID: 30518635; PubMed Central PMCID: PMC6278587.