



# Unraveling the impact of host genetics and factors on the urinary microbiome in a young population

Leying Zou,<sup>1</sup> Zhe Zhang,<sup>1</sup> Junhong Chen,<sup>1</sup> Ruijin Guo,<sup>2</sup> Xin Tong,<sup>2</sup> Yanmei Ju,<sup>2,3</sup> Haorong Lu,<sup>4</sup> Huanming Yang,<sup>2,5</sup> Jian Wang,<sup>2,5</sup> Yang Zong,<sup>2</sup> Xun Xu,<sup>2</sup> Xin Jin,<sup>2</sup> Liang Xiao,<sup>2,6</sup> Huijue Jia,<sup>7</sup> Tao Zhang,<sup>1,8</sup> Xiaomin Liu<sup>1,8</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 16.

**ABSTRACT** The significance of the urinary microbiome in maintaining health and contributing to disease development is increasingly recognized. However, a comprehensive understanding of this microbiome and its influencing factors remains elusive. Utilizing whole metagenomic and whole-genome sequencing, along with detailed metadata, we characterized the urinary microbiome and its influencing factors in a cohort of 1,579 Chinese individuals. Our findings unveil the distinctiveness of the urinary microbiome from other four body sites, delineating five unique urotypes dominated by *Gardnerella vaginalis*, *Sphingobium fluviale*, *Lactobacillus iners*, *Variovorax* sp. PDC80, and *Acinetobacter junii*, respectively. We identified 108 host factors significantly influencing the urinary microbiome, collectively explaining 12.92% of the variance in microbial composition. Notably, gender-related factors, including sex hormones, emerged as key determinants in defining urotype groups, microbial composition and pathways, with the urinary microbiome exhibiting strong predictive ability for gender (area under the curve [AUC] = 0.843). Furthermore, we discovered 43 genome-wide significant associations between host genetic loci and specific urinary bacteria, *Acinetobacter* in particular, linked to eight host loci ( $P < 5 \times 10^{-8}$ ). These associations were also modulated by gender and sex hormone levels. In summary, our study provides novel insights into the impact of host genetics and other factors on the urinary microbiome, shedding light on its implications for host health and disease.

**IMPORTANCE** The urinary microbiome, essential to human health, reveals its unique qualities in our study of 1,579 Chinese individuals. We identified distinctive microbial profiles, or “urotypes,” and uncovered strong gender-related influences, particularly from sex hormones, on these microbial communities. Our research highlights significant genetic associations affecting specific urinary bacteria, indicating a deep interaction between our genetics and our microbiome. These insights not only enhance our understanding of the urinary microbiome’s role in health and disease but also open new pathways for personalized medical strategies, making our findings crucial for future diagnostic and therapeutic innovations. This work underscores the intricate relationship between our body’s biological processes and the microorganisms within, providing valuable knowledge for both scientific and medical communities.

**KEYWORDS** urinary microbiome, metagenomics, host genetics, sex hormone

**M**ounting evidence has implicated the presence of a microbiome community within the urinary tract, commonly referred to as urinary microbiome or urobiome (1, 2). This discovery has been substantiated through high-throughput 16S rRNA/metagenomic sequencing, as well as enhanced urine culture techniques across multiple studies (3–6). In recent years, the urinary microbiota has garnered considerable attention owing to its potential implications for urinary tract health and diseases. For instance, studies

**Editor** Maria Gloria Dominguez Bello, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA

Address correspondence to Xiaomin Liu, liuxiaomin@genomics.cn, Tao Zhang, tao.zhang@genomics.cn, or Huijue Jia, huijue\_jia@fudan.edu.cn.

Leying Zou, Zhe Zhang, and Junhong Chen contributed equally to this article. The order of co-first authors is determined by the extent of their contributions and mutual agreement among the authors.

The authors declare no conflict of interest.

**Received** 10 September 2024

**Accepted** 16 October 2024

**Published** 8 November 2024

Copyright © 2024 Zou et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

have shown that the urinary microbiome differs significantly in urinary incontinence patients with increased *Gardnerella* and decreased *Lactobacillus* compared to healthy controls (3, 7, 8). The urinary microbiome's composition can also influence susceptibility to urinary tract infections (UTIs) (9–11). It has been observed that the postoperative UTI risk is associated with enrichment of a mixture of uropathogens, as well as depletion of *Lactobacillus* in the preoperative urinary microbiome (12). Furthermore, the urinary microbiome was shown to exhibit distinct signatures between bladder cancer patients and healthy controls, such as increased abundance of *Fusobacterium* in the patient group (13, 14). Such observations, coupled with the fact that intravesical BCG administration as the gold-standard adjuvant immunotherapy for treating high-risk non-muscle-invasive bladder cancer (15), suggest a link between urinary microbiome and cancer development and progression potentially through mechanisms involving inflammation and modulation of the immune response.

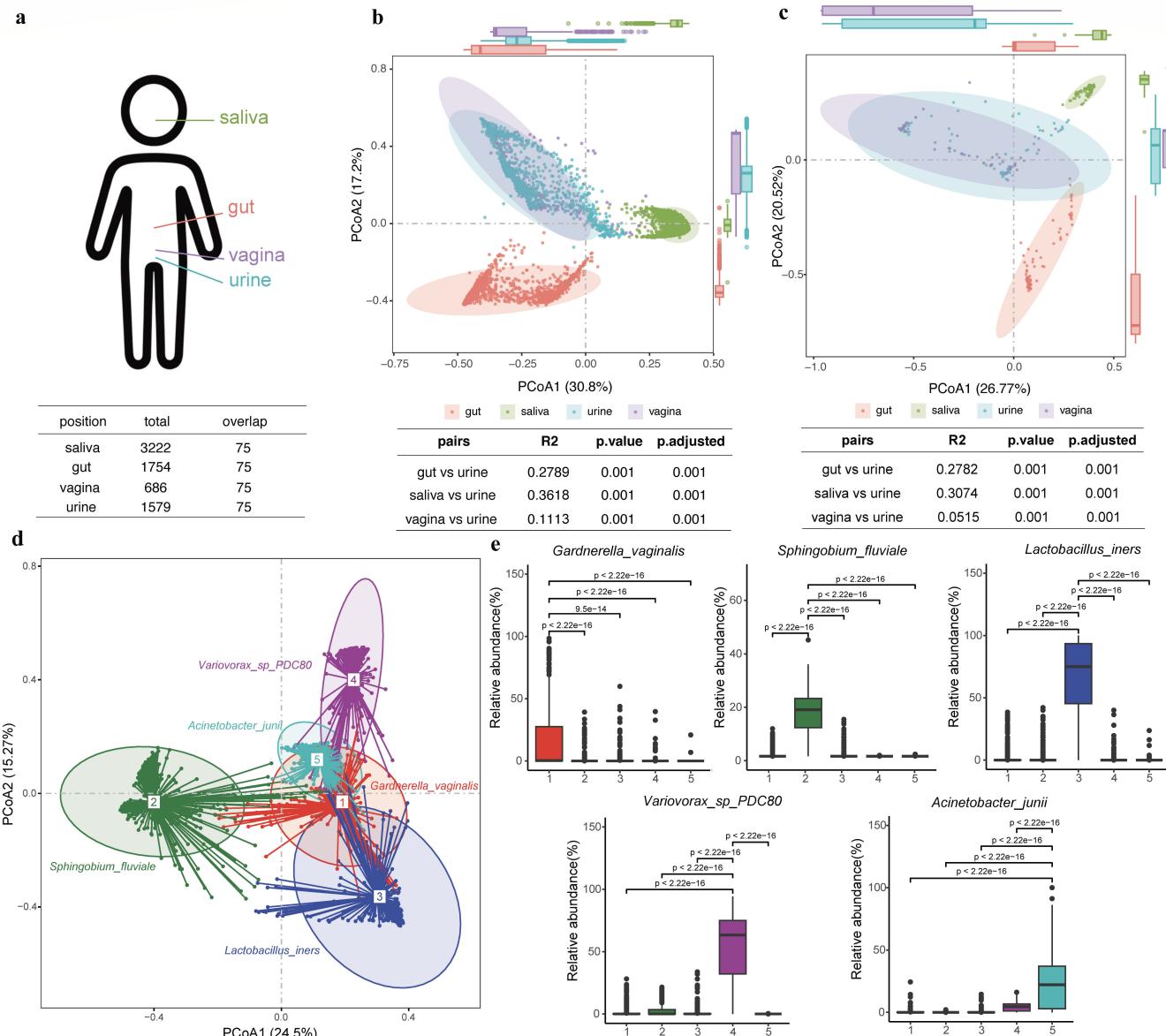
Thus, better understanding of the urinary microbiome and elucidating the factors that influence its composition and diversity are essential for developing effective diagnostic and therapeutic interventions. Previous studies have identified certain host factors, such as sex and age, that may influence composition of the urinary microbiota (16, 17). A recent study in mainly postmenopausal women has suggested a role for host genetics in microbial variations among individuals (18). However, these studies have predominantly utilized 16S rRNA sequencing that had relatively low resolution and allowed for the identification and quantification of bacterial taxa at the genus level. Furthermore, existing research has primarily focused on elderly or middle-aged individuals, in which the microbiome may be largely influenced by the environmental factors in the aging process. Thus, there is a gap in knowledge concerning urinary microbiota in young individuals, which our study aims to address. Moreover, there is a necessity to explore additional potential host factors, such as blood metabolites, urine factors, lifestyle, etc., as they may also impact the urinary microbiota. Current research in this area remains limited and calls for more in-depth exploration.

Here in this study, leveraging shotgun metagenomic sequencing and comprehensive metadata from 1,579 Chinese individuals from the 4D-SZ cohort (19–24), which primarily comprised young adults around 30 years old, we conducted a thorough investigation to enhance our understanding of the urinary microbiota and its influencing factors. Initially, we characterized the urinary microbiome by analyzing its microbial composition and comparing it to microbial communities in other body sites. We then systematically investigated its influencing factors among 386 traits, encompassing sex, age, body mass index (BMI), diets, lifestyles, as well as blood and urinary measurements. Moreover, by performing the so-far first urinary metagenome-genome-wide association study (M-GWAS) in a subset of 687 individuals with whole-genome sequencing data available, we endeavored to unravel the intricate interplay between host genetics and the urinary microbiome.

## RESULTS

### Urinary microbiome: a unique microecology distinct from other body sites

We collected urine samples of 1,579 individuals, with 66.12% being females and an average age of 29 years old. Utilizing shotgun metagenomic sequencing, we generated an average of  $19.18 \pm 7.90$  GB data per sample (Table S1; Fig. S1). These samples are from the 4D-SZ cohort (19–24), with multi-omics data collected including human whole genome, whole metagenomes across multiple body sites, blood metabolites, and detailed questionnaires. Because we have metagenomic data from not only the urinary tract but also the other body sites (Fig. 1a), including the gut ( $n = 1,754$ ), saliva ( $n = 3,222$ ), and reproductive tract ( $n = 686$ ), it prompted us to compare the urinary microbiome with the microbial communities inhabiting these distinct body sites. Permutational multivariate analysis of variance (PERMANOVA) revealed that the urinary microbiome is significantly distinct from the other examined microbiomes. It showed a much lower  $\alpha$ -diversity compared to the rich and diverse microbiome harbored in



**FIG 1** Comparative analysis of microbial communities across human body sites and characterizing the urinary microbiome. (a) An illustrative diagram presenting the total and shared sample counts of microbial communities from four body sites: saliva, gut, urine, and the reproductive tract. (b) A scatterplot illustrating the genus-level PCoA distribution of the total microbial samples from saliva, gut, urine, and reproductive tract, along with PERMANOVA pairwise difference values. (c) A scatterplot showing the genus-level PCoA distribution of 75 shared samples among saliva, gut, urine, and reproductive tract, with PERMANOVA pairwise difference values. (d) A scatterplot reflecting the species-level PCoA distribution of the urinary microbes, the urotypes classification, and the dominant bacteria present in each urotype. (e) Boxplots displaying the average relative abundance of the five dominant bacteria among individuals with five different urotypes, along with t-test differential analysis.

the gut and the oral (25), with the highest disparity observed in the phylum level with the gut microbiome ( $R^2 = 0.455$ , higher  $R^2$  values corresponding to a larger difference between the compared sites,  $P < 0.001$ ; Fig. S2) and in the genus level with the saliva microbiome ( $R^2 = 0.361$ ,  $P < 0.001$ ; Fig. 1b). In comparison to the vaginal microbiome, which is known to characterize a very low  $\alpha$ -diversity (25), the urinary microbiome exhibited the smallest but significant dissimilarity at both phylum ( $R^2 = 0.214$ ,  $P < 0.001$ ) and genus ( $R^2 = 0.111$ ,  $P < 0.001$ ) levels (Fig. S2; Fig. 1b), albeit with a significantly higher  $\alpha$ -diversity (Fig. S3).

Further analysis was conducted on metagenomic data from 75 individuals to corroborate the above findings, wherein data from all four body sites were available. Consistent with the broader data set, the urinary microbiome exhibited the most pronounced differences from the gut microbiome at the phylum level ( $R^2 = 0.299$ ; Fig. S4) and from the salivary microbiome at the genus level ( $R^2 = 0.307$ ; Fig. 1c) while displaying relatively smaller disparities with the vaginal microbiome ( $R^2 = 0.11$  at the phylum level and  $R^2 = 0.051$  at the genus level; Fig. 1c; Fig. S4). Importantly, these variations remained statistically significant (adjusted  $P < 0.001$ ). Furthermore, we conducted a fast expectation-maximization for microbial source tracking (FEAST [26]) analysis to estimate the potential attributions for the urinary microbiome from the other body sites in these 75 individuals. The results showed the vaginal microbiome was predicted as the main contributor accounting for an average of 60.45% of the urinary microbiome, while saliva (0.48%) and gut (0.27%) microbiomes were estimated to have limited contribution in these young adults (Fig. S5).

It is evident that these distinctions arise from variations in genus composition across different body sites (Table S2). For example, *Lactobacillus* was the most abundant genus found in both urinary (an average relative abundance of 17.6%) and vaginal samples (58.9%), while its average relative abundance in saliva and gut was less than 0.13%. The second and third most prevalent genera in urine were *Variovorax* and *Acinetobacter*, with mean relative abundances of 10.6% and 9.4%, respectively; however, their average relative abundances in the other examined sites were less than 0.03%. *Prevotella* is found to be enriched in the saliva with an average relative abundance of 29.1%, while it is less abundant in the gut at 13.4%, followed by urine at 6.7% and vagina at 3.6%. The distinctive microbial compositions and species' diversities across different human body sites resulted in marked dissimilarities in the microbiota across these regions.

In summary, our findings underscore the uniqueness of the urinary microbiota in comparison to other anatomical regions such as the gut, saliva, and vagina. This distinct microbial ecosystem warrants further in-depth investigation.

### Characteristics of the urinary microbiome

Because the urinary microbiome represents a distinctive microecology, we further characterized its community composition in detail. First, we examined the top 10 most abundant genera and species in the cohort (Fig. S6). At the genus level, *Lactobacillus* had the highest relative abundance, followed by *Variovorax*, *Acinetobacter*, *Prevotella*, and *Sphingobium*. At the species level, the top five bacteria with the highest relative abundances, in descending order, were *Lactobacillus iners*, *Variovorax* sp. PDC80, *Sphingobium fluviale*, *Lactobacillus crispatus*, and *Gardnerella vaginalis*. Notably, we observed high Pearson correlations between the relative abundances of bacteria commonly found in the vagina, such as *Lactobacillus*, *Gardnerella*, *Prevotella*, and *Ureaplasma*, in the urinary and vaginal microbiomes (Pearson's  $r > 0.5$ ).

Next, using the partitioning around medoids (PAM) clustering method (27), the urinary microbial communities were clustered into five distinct clusters (referred to as "urotypes" analogous to the "enterotypes" [28]). By applying linear discriminant analysis (LDA) effect size (LEfSe) (29) for detecting species discriminating each urotype from the rest and ranking them by LDA score, we determined the dominant species for each urotype. They are *G. vaginalis* (accounting for 23.10% of the samples), *S. fluviale* (34.11%), *L. iners* (15.76%), *Variovorax* sp. PDC80 (16.58%), and *Acinetobacter junii* (10.38%), respectively (Fig. 1d). Consistent with the clustering results, each urotype showed a significant enrichment of the respective dominant bacteria (Fig. 1e).

While the dominant species characterized the highest LDA score for each urotype, there were urotype-characteristic species identified, especially in clusters 1, 2, and 5 (Fig. S7; Table S3). For cluster 1 (*G. vaginalis* urotype), the second-ranked species *L. crispatus* followed the dominant bacterium *G. vaginalis* so close in LDA score, just like their close relative abundances with an average of 3.91% versus 3.92%. The individuals characterizing this urotype also harbored significantly higher levels of species belonging

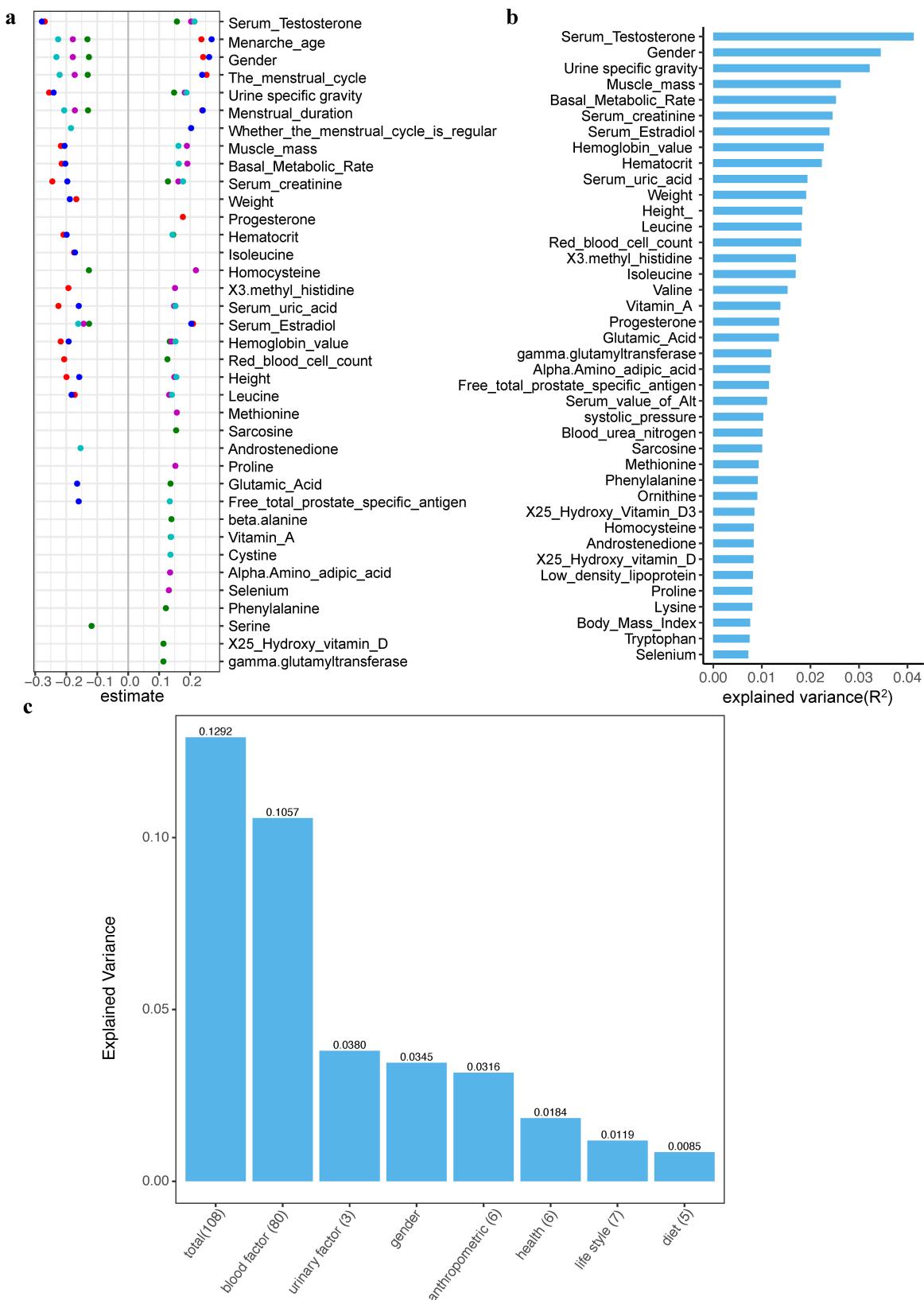
to the genus *Prevotella*, including *Prevotella bivia* (3.67%), *Prevotella timonensis* (3.43%), *Prevotella disiens* (3.08%), *Prevotella colorans* (2.84%), and *Prevotella corporis* (2.70%). For cluster 2 (*S. fluviale* uotype), species such as *Sphingomonas ursincola* (3.76%), *Novosphingobium* sp. AAP93 (3.62%), *Erythrobacter donghaensis* (3.59%), and *Acinetobacter johnsonii* (3.56%) were also highly represented. For cluster 3 (*L. iners* uotype), *Lactobacillus jensenii* ranked second with an LDA score of 2.7 and belonged to the same genus *Lactobacillus* as the dominant species. In cluster 4 (*Variovorax* sp. PDC80 uotype), *Bradyrhizobium* sp. CCH1 B1 also showed a high proportion with an LDA score of 3.76. Finally, in cluster 5 (*A. junii* uotype), *Herbaspirillum aquaticum* (3.50%), *Cutibacterium acnes* (3.32%), and *Moraxella osloensis* (3.31%) were also most abundant and significantly enriched compared to the other clusters.

Microbial pathways were also investigated (Fig. S8; Table S4). S-adenosyl-L-methionine salvage I (PWY-6151), sucrose biosynthesis II (PWY-7238), and guanosine ribonucleotides *de novo* biosynthesis (PWY-7221) were enriched in cluster 1 (*G. vaginalis* as the dominant bacteria). Consistently, *G. vaginalis* contributed to PWY-6151 (30), which has been previously linked to anti-proliferative, pro-apoptotic, and anti-metastatic processes in several types of cancer (31). Similarly, *L. iners* and its contributed pathway pyruvate fermentation to acetate and lactate (PWY-5100, P41-PWY) (30) were both most enriched in individuals of cluster 3. PWY-5136: fatty acid beta-oxidation II (plant peroxisome), PWY-7663: gondoate biosynthesis (anaerobic), and PWY-4984: urea cycle were mainly enriched pathways in clusters 1, 4, and 5, respectively. We found bacteria and their contributing pathways were consistently enriched in each uotype cluster.

### Gender-related factors determined uotypes and urinary microbial communities

To understand the factors influencing uotypes, we calculated their correlations with host factors. Notably, gender-related factors emerged as the primary determinants of these uotypes. Menstrual cycle, serum testosterone, serum estradiol, serum creatinine, proportion of urine, menstrual duration, menarche age, hemoglobin, and gender were significantly associated with all five uotypes (Fig. 2a; Table S5). Remarkably, gender distinctly differentiated the uotype groups. Females exhibited higher relative abundances of *G. vaginalis* (uotype 1) and *L. iners* (uotype 3), aligning their urinary microbiome predominantly with these two uotypes. In contrast, males demonstrated higher abundances of *S. fluviale* (uotype 2), *Variovorax* sp. PDC80 (uotype 4), and *A. junii* (uotype 5), characterizing their urinary microbial communities with these three respective uotypes. Additionally, the gender-biased uotypes displayed significant correlations with gender-dependent metadata. For instance, uotypes 4 and 5, predominantly found in males, correlate with higher levels of serum testosterone, urine-specific gravity, muscle mass, basal metabolic rate, serum creatinine, and serum uric acid—clinical and biochemical parameters that are typically elevated in males compared to females. Conversely, uotypes 1 and 3, more common in females, show a significant correlation with increased serum estradiol levels. Intriguingly, the gut microbiota also display sex differences, correlating significantly with gonadal steroids such as serum testosterone (32). Additionally, it is well established that estrogen can promote vaginal *Lactobacillus* level by accumulated glycogen production in the vaginal squamous epithelial cells (33). It might be reasonable to speculate that the relatively higher levels of estrogen in women, which are partly excreted through urine, may similarly also be in relation with enhanced growth of *Lactobacillus* within the urinary tract.

We further delved into host factors that contributed to the overall composition of the urinary microbiome and identified a total of 137 significant influencing factors ( $P$  adjusted  $< 0.05$ ; Table S6). Serum testosterone, gender, urine-specific gravity, muscle mass, basal metabolic rate, serum creatine, and estradiol were among the top influencing factors associated with  $\beta$ -diversity of the urinary microbiome (Fig. 2b). Due to some factors that were highly correlated, we calculated the Spearman correlations among the 137 host factors and removed 29 collineated factors (Spearman's  $r > 0.6 \parallel r < -0.6$ ). The



**FIG 2** Impact of host phenotypes on urinary microbiota. (a) This scatterplot presents the distribution of Pearson correlation estimate values between host phenotypes and the five urotypes, highlighting the top 20 significant associations for each urotype, with colors corresponding to the urine classification results in Figure 1d. Each phenotype is ranked from high to low according to the average of Pearson correlation estimate value for each urotype. (b) A bar chart of explained variance ( $R^2$ ) for individual phenotypes. (Continued on next page)

## Fig 2 (Continued)

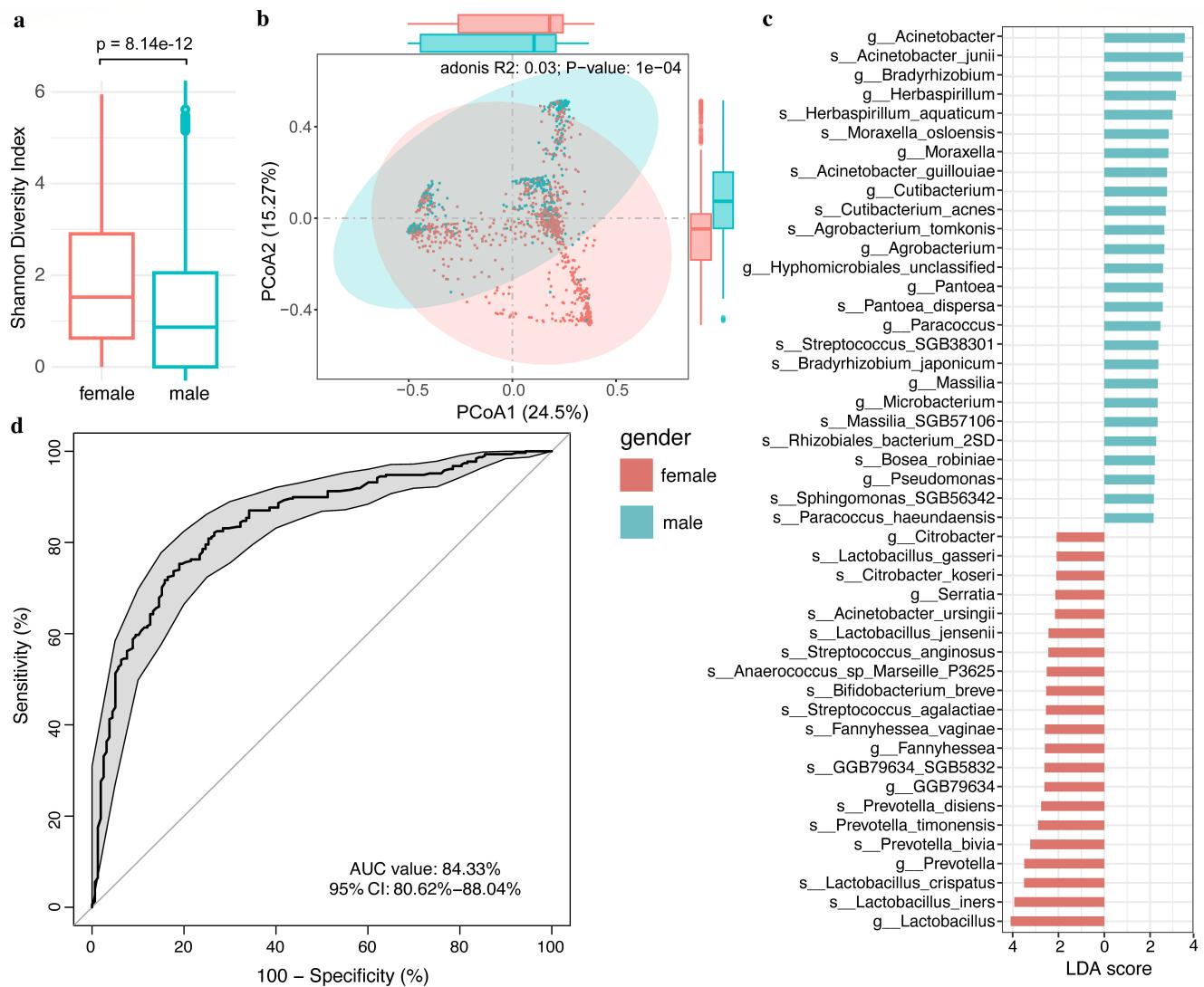
showing the top 40 significant host phenotypes sorted by their explained variance ( $R^2$  values) in the  $\beta$ -diversity of urinary microbiota. (c) A categorized bar chart displaying the explained variance of significant host phenotypic factors for the urinary microbiota  $\beta$ -diversity, including 108 total independent host factors, 80 blood factors, 3 urinary factors, gender, 6 anthropometric factors, 6 health factors, 7 lifestyle factors, and 5 diet-related factors.

remaining 108 independent and significant covariates collectively explained 12.92% of the variance of microbial composition (Fig. 2c), which encompassed 80 blood factors (explained  $R^2 = 10.57\%$ ), 3 urinary factors (3.80%), gender (3.45%), 6 anthropometric measurements (3.16%), 6 health-related factors (1.84%), 7 lifestyles (1.19%), and 5 dietary factors (0.85%). Among the 80 blood factors, hematocrit (2.3%), serum uric acid (2.0%), 3-methyl-histidine (1.8%), and isoleucine (1.8%) exhibited substantial impacts on the urinary microbiome. The collective urinary factors ranked as the second-most significant influence among the seven categories, with urine-specific gravity demonstrating the highest explanatory power (3.28%). Among the six anthropometric factors, height (1.83%) and systolic pressure (1.03%) had explanatory powers that exceeded BMI (0.76%). Within the health-related factors, the test result of the free total prostate specific antigen had the most significant impact on the urinary microbiome (1.14%). In contrast, dietary factors had a comparatively lower effect, with the highest influencing factor being fat meat favorites, which had an explanatory power of only 0.25% (Table S6).

In addition to individual urotypes and microbial composition, we also identified 69 host factors significantly associated with 115 microbial pathways ( $P$  adjusted  $< 0.05$ ; Table S7). Consistent with the findings for microbial composition, urine-specific gravity, serum testosterone, and gender were the top three strongest host factors correlated with various of microbial pathways. They were all linked to pathways such as HISTSYN-PWY: L-histidine biosynthesis, HEME-BIOSYNTHESIS-II: heme b biosynthesis I (aerobic), HEME-BIOSYNTHESIS-II-1: heme b biosynthesis V (aerobic), PWY-5136: fatty acid beta-oxidation II (plant peroxisome), and TRPSYN-PWY: L-tryptophan biosynthesis. Testosterone administration has been reported to reduce the urinary excretion of histamine in female mice (34) and increased plasma tryptophan levels in women (35). Further research is warranted to elucidate the precise mechanisms through which these factors interact with the microbial pathways.

Motivated by the insights that gender and gender-related factors are primary influencers of the urinary microbiome, we proceeded to investigate differences in the microbial community between genders in detail. We first observed significantly higher  $\alpha$ -diversity, as measured by the Shannon index, in the urinary microbiota of females than in males ( $P < 8.14 \times 10^{-12}$ , Fig. 3a). The microbiota composition also significantly differed between males and females ( $R^2 = 0.03$ ,  $P < 1 \times 10^{-4}$ , Fig. 3b). Furthermore, we identified 47 sex-differentiated microbial features including 18 genus and 29 species, by using LEfSe analysis (Fig. 3c). Females were characterized by higher abundances of genera *Lactobacillus*, *Prevotella*, *Fannhyessea*, *Serratia*, and *Citrobacter*. However, males exhibited higher abundances in species belonging to *Acinetobacter*, *Bradyrhizobium*, *Herbaspirillum*, *Moraxella*, *Cutibacterium*, and other genera. These results were in line with the Spearman correlations between bacterial profiles and serum testosterone (Fig. S9), suggesting the important effects of sex and sex hormones on shaping the urinary microbiota.

Given the striking gender-related difference in the microbial community, we employed a random-forest-based machine learning approach to assess the predictive capability of urinary microbiota for gender classification. We used the 47 sex-differentiated microbial features for random forest modeling, with 70% of the samples as training sets and 30% as test sets. The predictive accuracy demonstrated by the urinary microbiota in determining gender was remarkably high, with an area under the curve (AUC) value of approximately 0.843 (Fig. 3d). Notably, the representative species of urotypes including *L. iners* and *A. junii* were among the top 10 species that contributed most significantly to gender differentiation according to Gini importance (Table S8). These findings highlight the intricate relationship between microbial composition and host



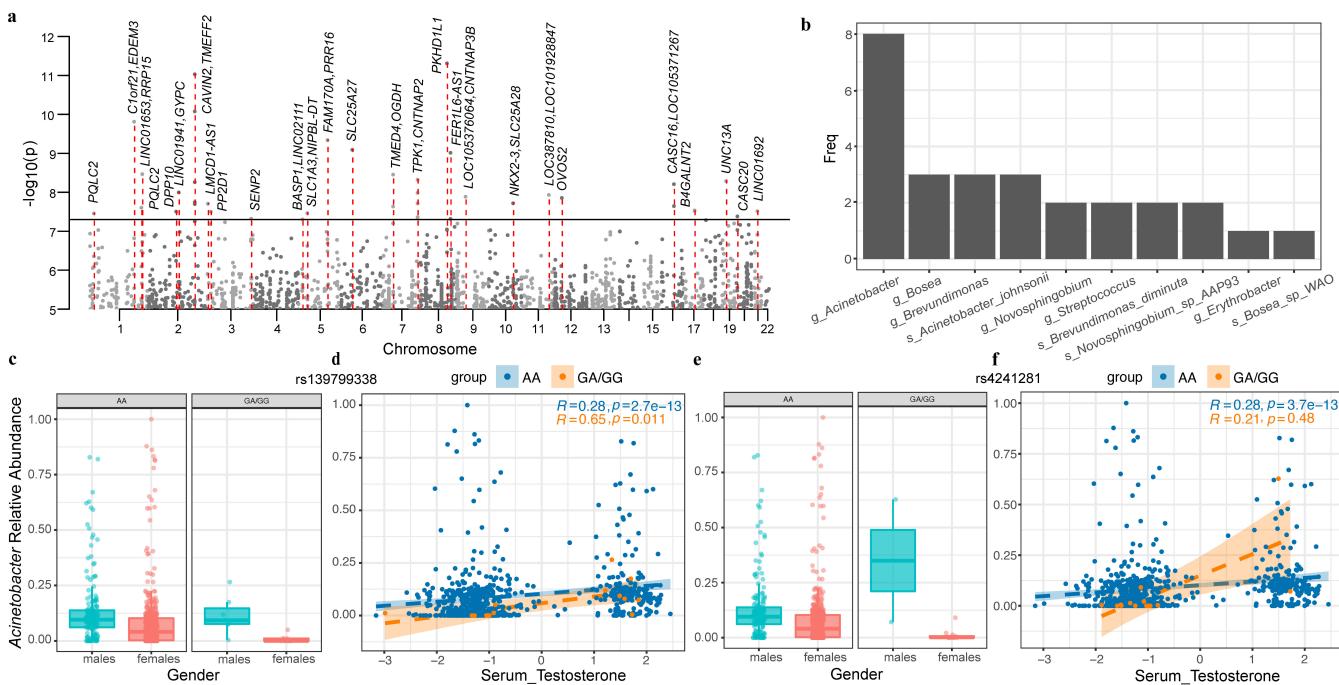
**FIG 3** The urinary microbiome differed between genders and effectively distinguished gender. (a) Boxplot of  $\alpha$ -diversity (Shannon index) of the urinary microbiome between males and females. (b)  $\beta$ -Diversity of the urinary microbiome using PCoA, as colored by males (cyan) and females (red). The first and second axes, PCoA1 and PCoA2, explaining the highest variance, are shown. Dashed ellipses represent the 95% confidence level of PERMANOVA test. Explained variance ( $R^2$ ) and  $P$  value are also shown. (c) The sex differential urinary microbiota, as sorted by LDA scores in the LEfSe analysis. (d) A ROC curve demonstrating the predictive power of the urinary microbiota for gender using a random forest model. The analysis comprises 520 male samples and 1,044 female samples.

biological characteristics, suggesting that the urinary microbiome's distinct makeup is reflective of and influenced by the host's biological sex.

### Host genetics' influences on the urinary microbiota

Out of a total of 1,579 individuals, 687 had host whole-genome sequencing data available, which enabled us to investigate the influence of the host genome on the urinary microbiome. To achieve this, we conducted an M-GWAS on 106 urinary taxa present in over 10% individuals (Table S9). After adjusting for covariates such as gender, age, BMI, sequencing read counts, and the top 10 host principal components (PCs), we identified 43 genome-wide significant associations involving 27 host genetic loci and 10 urinary taxa ( $P < 5 \times 10^{-8}$ , Fig. 4a; Table S10).

Of all the microbial taxa examined at the genus and species levels, the genus *Acinetobacter* was associated with most host genetic loci ( $n = 8$ ; Fig. 4b). The strongest association was observed for SNP rs139799338, located in the intronic region of the



**FIG 4** Influences of host genetics and their interactions with gender/sex hormones on urinary microbiota. (a) A Manhattan plot showing the genetic variants associated with urine microbial taxa ( $n = 106$ ), with genome-wide significance denoted by horizontal black lines ( $P = 5 \times 10^{-8}$ ) and gene names annotated for 27 significant loci. (b) A bar chart displaying the urinary microbes associated with the significant 27 loci and the number of associations. (c) Boxplots illustrating the average relative abundance of the genus *Acinetobacter* among individuals with different genotypes (AA vs GA/GG) at the top locus rs139799338, as well as stratified by male and female samples. (d) A scatterplot depicting the association between serum testosterone and the genus *Acinetobacter* relative abundance among individuals with different genotypes (AA vs GA/GG) at the top locus rs139799338. Spearman correlation's R and P values are also listed. (e) Boxplots showing the average relative abundance of the genus *Acinetobacter* among individuals with different genotypes (AA vs GA/GG) at the top locus rs4241281, as well as stratified by male and female samples. (f) A scatterplot illustrating the association between serum testosterone and the genus *Acinetobacter* relative abundance among individuals with different genotypes (AA vs GA/GG) at the top locus rs4241281. Spearman correlation's R and P values are also listed.

*PKHD1L1* gene ( $P = 4.82 \times 10^{-12}$ ), negatively associated with the relative abundance of the *Acinetobacter* genus (Fig. 4a). *PKHD1L1* has been found to be significantly elevated in the urine of clear cell renal cell carcinoma patients (36). It was also reported to primarily express in both infant and adult human kidneys and is associated with autosomal recessive polycystic kidney disease (37). The SNP rs139799338 demonstrated associations with various metabolic traits, including the tumor marker carbohydrate antigen 199 ( $P = 0.006$ ), glycine ( $P = 0.01$ ), vitamin A ( $P = 0.01$ ), beta-alanine ( $P = 0.03$ ), and cobalt ( $P = 0.047$ ), as revealed by the genome-wide association study (GWAS) analysis of host genetic variants and metabolic traits in this cohort (Fig. S10; Table S11).

The second-most significant correlation identified was with rs4241281, located in the intergenic region of genes *CAVIN2* and *TMEFF2*, which was found to have a negative association with the genus *Acinetobacter* ( $P = 9.13 \times 10^{-12}$ , Fig. 4a; Table S10). This genetic locus was also correlated with the relative abundance of the genus *Bosea*, ranking as the third strongest association. Additionally, it was also linked to the abundance of the species *A. johnsonii* belonging to the genus *Acinetobacter* ( $P = 1.75 \times 10^{-8}$ ). *CAVIN2* encodes Caveolae Associated Protein 2 in the kidney, and its removal leads to caveolae loss, while overexpression results in caveolae deformation and membrane tubulation (38). *TMEFF2* has been identified as an epigenetic biomarker for bladder cancer (39), showing significant associations with increasing tumor grade and stage. Notably, SNP rs4241281 in the *CAVIN2-TMEFF2* was also correlated with several blood and urinary factors, including concentrations of mercury ( $P = 0.007$ ), alpha amino adipic acid ( $P = 0.015$ ), albumin ( $P = 0.035$ ), body fat mass and rate ( $P = 0.034 \sim 0.017$ ), and urine pH ( $P = 0.047$ , Table S11). Individuals with the rs185971980 SNP located in the intergenic

region of *FAM170A* and *PRR16* exhibited lower abundance of *A. johnsonii* in the urine ( $P = 4.46 \times 10^{-10}$ ). Additionally, the rs185971980 was correlated with levels of mean hemoglobin content, mean erythrocyte hemoglobin concentration, average red blood cell (RBC) volume, and blood iron. Similarly, individuals with the rs117738684 SNP had lower abundance of *A. johnsonii* ( $P = 3.44 \times 10^{-9}$ ). The rs117738684 was also correlated with levels of blood testosterone, fasting sugar, methionine, and albumin. Moreover, the SNP rs9975224 within *LINC01692* was associated with the abundance of not only the species *Novosphingobium* sp. AAP93 but also the urine pH ( $\beta = 0.087, P = 9.3 \times 10^{-4}$ ). SNPs within *SENP2* was associated with the abundance of not only the genus *Acinetobacter* but also the urine-specific gravity ( $\beta = -0.087, P = 1.8 \times 10^{-4}$ , Table S11). These results suggested host genes may influence the urinary microbiota by regulating the blood metabolites and urinary environment.

Gender and sex hormones have demonstrated significant impacts on the urinary microbiome, prompting an exploration of gender–gene interactions using the top two loci associated with specific taxa (Fig. 4c through f). An intriguing observation surfaced in our study, revealing that males with heightened testosterone levels displayed an increased abundance of *Acinetobacter* compared to females, regardless of their genetic makeup (genotypes AA or GA/GG at SNP rs139799338). Additionally, carriers of genotype AA exhibited a higher abundance of *Acinetobacter* compared to those carriers of genotype GA/GG in females, while no apparent distinctions were noted among males with varying genotypes (Fig. 4c). Similarly, an interactive effect between rs139799338 genotypes and serum testosterone levels on *Acinetobacter* was observed. Specifically, a stronger correlation of serum testosterone level and *Acinetobacter* abundance was noted in individuals with genotype GA/GG ( $R = 0.65, P = 0.011$ ) compared to individuals with genotype AA ( $R = 0.28, P = 2.7 \times 10^{-13}$ , Fig. 4d). The gender–gene interaction was also evident for another SNP, rs4241281, linked to *Acinetobacter*, with males exhibiting a higher average relative abundance than females (Fig. 4e). In individuals with the rs4241281 AA allele, serum testosterone demonstrated a strong correlation with *Acinetobacter* ( $R = 0.28, P = 3.7 \times 10^{-13}$ , Fig. 4e and f).

### Potential causal relationships between urinary microbiome and diseases

Finally, to explore the potential causal relationships between urinary microbes and diseases by leveraging microbiome-associated host genetic variations, we performed a two-sample Mendelian randomization (MR) analysis. After adjusting for 42 diseases from the Biobank Japan (BBJ) cohort (40), we identified seven suggestive causal relationships ( $P \leq 1.19 \times 10^{-3} = 0.05/42$ , Table S12; Fig. S11). The strongest MR evidence suggested a link between type 2 diabetes (T2D) and a decreased abundance of *Streptococcus* SGB38301 ( $\beta = -0.076, P = 2.83 \times 10^{-4}$ ). *Paracoccus haematequi* was causally correlated with a decreased risk of T2D ( $\beta = -0.013, P = 5.12 \times 10^{-4}$ ), while the males-enriched *Variovorax* sp. PDC80 linked to an increased T2D risk ( $\beta = 0.017, P = 8.89 \times 10^{-4}$ ). The genus *Herbaspirillum* potentially increased the risk of atopic dermatitis ( $\beta = 0.064, P = 4.41 \times 10^{-4}$ ). *Bosea robiniae* potentially increased the risk of prostate cancer ( $\beta = 0.058, P = 4.65 \times 10^{-4}$ ). Keloid was associated with a lower *Bosea* ( $\beta = -0.061, P = 7.64 \times 10^{-4}$ ). These inferred relationships underscore a potential connection between urinary microbes and autoimmune diseases, warranting validation in future studies.

### DISCUSSION

This study represents the initial profiling of the urinary microbiome in young adults in comparison to well-studied human microbiomes from diverse body sites such as the gut, oral cavity, and reproductive tract. While previous studies have also compared the microbial composition across various body sites (18), they relied on databases from different populations and generally involved smaller sample sizes. In contrast, our research provides a more cohesive and comprehensive analysis by utilizing a uniform cohort.

The urinary microbiome was found to possess unique characteristics compared to other sites, evident in the composition of microbial species, as well as  $\alpha$ -and  $\beta$ -diversity differences. The urinary microbiome exhibited significant dissimilarity from other sites, as demonstrated by PCoA analysis in both the entire sample set and the subset of 75 individuals with data from different sites. The difference in microbial communities across body sites was primarily driven by specific taxa. For instance, the gut microbiome was predominantly composed of the genera *Bacteroides* and *Prevotella*, whereas the saliva microbiome showed a gradient of abundance among *Prevotella*, *Neisseria*, and *Haemophilus*. The urinary microbiome was predominantly composed of genera such as *Lactobacillus*, *Variovorax*, and *Acinetobacter*, while the vaginal microbiome exhibited more abundant *Lactobacillus* and *Gardnerella*. This highlights the significant role that specific microbial taxa play in characterizing the microbial landscape of different body sites, indicating distinct ecological niches and potentially different functional roles within the human body. Despite the differences, urine and vagina harbored several predominant bacteria in common, supporting the relatively closer relatedness between them (41) compared to other physically distant sites.

We identified five uotypes and their respective dominant bacteria. A previous study extracted microbiota from encrustations on ureteral stents in patients without urinary tract infections and conducted uotype analysis (42). Their findings, similar to ours, identified types dominated by microbes such as *Lactobacillus*, *G. vaginalis*, *Achromobacter*, *Corynebacterium*, and *Streptococcus* (Table S3). However, they identified extra bacteria like *Actinomycetales*, *Enterobacteriales*, and *Staphylococcus*, which were less abundant in this cohort. These differences probably have stemmed from the distinct sources of the analyzed microbiome. Their study focused on bacteria from ureteral stents, on which nosocomial bacteria may form biofilm. They employed the 16S rRNA sequencing method, which differs from our metagenomic sequencing approach.

Although some studies indicated the urinary microbiome may be impacted by sex (43–46), they were performed in small samples and mainly based on the 16S rRNA data, resulting in limited sex-specific bacteria identification. In this study, we discovered extensive sex differences in the urinary microbiota, which could well predict sex (AUC = 0.843). We also observed a strong link between the urinary microbiome and sex hormone levels as well as some other sex-specific factors. UTIs are among the most common bacterial infections affecting our urinary system (47), with uropathogenic *Escherichia coli* being the primary causative agent (11). Among healthy women aged 18–39 years, *E. coli* was responsible for 80% of UTI cases (48). In this study, we found that *E. coli* exhibited a higher abundance in women (2.96%) than in men (0.86%). *E. coli* was also enriched in uotype 1, a female-dominated uotype characterized by *G. vaginalis*. This may help to explain why the UTI incidence was higher in females than in males (47, 49). These findings not only enrich our understanding of the gender-specific susceptibilities to certain urinary tract conditions or diseases but also open avenues for utilizing microbial profiles in gender-specific medical research and potential diagnostic applications.

Our study also marks the first detailed exploration of the impact of blood metabolites on the urinary microbiome. Serum hormones, particularly serum testosterone and serum estradiol, emerged as significant influencers of the urinary microbiome. Other metabolites, such as serum creatinine and serum uric acid, along with blood amino acids like leucine, isoleucine, valine, glutamic acid, and vitamin A, were also found to play crucial roles, highlighting the intricate relationship between various serum components and the composition of the urinary microbiome. Serum creatinine has been regarded as a biomarker of renal function for almost a century (50) and is widely utilized in clinical diagnostics. In our study, serum creatinine showed significant explanatory power for the  $\beta$ -diversity of urinary microbiota and different uotypes, suggesting that serum creatinine levels might influence kidney function through its effects on the composition of urinary microbiota. The potential causal relationship warrants further investigation. This link emphasizes the intricate connections between metabolic biomarkers

and microbiome composition, highlighting the importance of a holistic approach to understanding renal health and disease.

Limited studies investigated the effect of host genome on the urinary microbiome, except a previous study that reported considerable genetic contribution and identified heritable *Escherichia* ( $A = 0.165$ ) and *L. iners* ( $A = 0.177$ ) by analyzing midstream urine samples from 1,600 older females in the TwinsUK cohort (18). This study represents a pioneering effort in conducting a genome-wide association analysis of the urinary microbiome along with the host genome. We have successfully identified 27 significant genomic loci associated with 10 urinary microbes, demonstrating the influence of genetic factors on the composition of the urinary microbiome. The genus *Acinetobacter* and its species *A. johnsonii* showed the most link to host genetic loci. In addition to the urinary microbiome, *A. johnsonii* exhibited a genetic attribute to the skin microbiome (51). In this GWAS, we also observed an association between the genus *Acinetobacter* and the *MARCO* gene ( $\beta = -0.183$ ,  $P = 3.67 \times 10^{-6}$ ), which encoded macrophage receptor with collagenous structure and correlated with stress urinary incontinence (replication  $P = 0.003$ ) in a GWAS of European women (52). Moreover, we delved deeper into exploring the impact of interactions between significant loci genotypes, gender, and serum testosterone levels on specific urinary microbes. The insights gained from these analyses offer valuable information on the complex interplay between host genetics, hormonal levels, and microbial abundance.

## MATERIALS AND METHODS

### Study subjects

All the participants involved in this research belonged to the “4D-SZ” cohort, which encompassed the collection of blood, gut, saliva, vaginal, and urine samples, along with comprehensive metadata, for a multi-omics investigation as previously reported (19–24). In this study, 1,579 urine samples from the cohort were collected for whole metagenomic sequencing (Table S1). The detailed urine sampling, sequencing, and quality control design are described below. For comparison, we also included metagenomic samples from the gut ( $n = 1,754$ ), saliva ( $n = 3,222$ ), and reproductive tract ( $n = 686$ ), all from the 4D-SZ cohort. Feces were collected with an MGIeasy kit, and stool DNA was extracted in accordance with the MetaHIT protocol as described previously (53). For the salivary sample, we employed a 2x concentration of a stabilizing reagent kit and collected 2 mL of saliva. The DNA extraction from saliva samples was conducted using the MagPure Stool DNA KF Kit B (no. MD5115-02B). For reproductive tract samples, the extraction of genomic DNA was performed in accordance with the established protocol (54). The DNA concentrations from blood, stool, oral, and reproductive tract samples were estimated by Qubit (Invitrogen), and the library preparation and sequencing were conducted in accordance with the procedures described in previous studies (19–24).

### Urine metagenomic sampling, sequencing, and quality control

Midstream voided urine was self-collected using a 40-mL urine cup by 1,579 individuals from the 4D-SZ cohort. From the urine cup, 5 mL urine was flowed into a sterile tube containing 5 mL BGI stabilizing reagent (55) for the preservation of metagenome at room temperature and then stored at  $-80^{\circ}\text{C}$  within the day before subsequent metagenomic analysis.

The samples were thawed and transferred into a 15-mL centrifuge tube. Then, 1 mL TE buffer (1 M Tris-HCl and 0.1 M EDTA) for crystal dissolution (56) and 200  $\mu\text{L}$  50 mg/mL lysozyme were added before centrifugation at 8,000 rpm for 10 minutes. Then, DNA extraction was carried out with MagPure Stool DNA KF Kit B (MD5115, Magen) (57).

DNA extraction and library preparation were performed along with blank negative controls and mock community controls. Specifically, ZymoBIOMICS Microbial Community Standard (D6300) and ZymoBIOMICS Microbial Community DNA Standard (D6305) were

$1 \times 10^4$  times diluted using saline to simulate the low microbial load in urine samples and were divided into aliquots (1 mL per tube) for storage at  $-80^{\circ}\text{C}$  before using as extraction and library preparation controls, respectively. During extraction, blank negative control and extraction control (D6300 diluted) were added for each 96-well plate and processed with the same procedure as for the urine samples. Similarly, separate blank negative control and library preparation control (D6305 diluted) were also added for each 96-well plate during library preparation and processed with the same procedure as for the urine samples. Notably, none of the negative controls for both extraction and library preparation passed the library preparation stage for PCR concentration lower than the minimum criteria of 1 ng/ $\mu\text{L}$ . In total, 33 extraction mock controls and 43 library mock controls generated metagenomic sequencing data.

### Filtering and removal of possible contaminants

As low biomass microbiome tends to be more easily influenced by contaminants and cross-talk, multiple contamination control steps were further applied to the sequencing data.

1. Step 1. For all the samples in the entire cohort, species were considered as possible contaminants and were collectively removed if a species meets one of the following criteria:
  - a. Relative abundance significantly correlated with PCR concentration (Spearman test,  $P$  value  $< 0.05$ )
  - b.  $\text{p.freq} < 0.3$  or occurrence in  $< 4$  samples. The value of  $\text{p.freq}$  was calculated by “decontam” package (58) based on relative abundance and PCR concentration
  - c. Detection in the 76 mock controls but other than the 10 expected species comprising the mock community with  $> 1 \times 10^{-4}$  mean relative abundance
  - d. Reads count less than 1,000.
2. Step 2. For all the samples in each specific 96-well plate, remaining species other than the 10 expected species detected in the respective mock control with  $> 1 \times 10^{-4}$  relative abundance were considered as contaminants specific to that 96-well plate and were removed.
3. Step 3. Non-bacterial microbes were filtered out, and only bacterial taxa and renormalization were retained.

During this process, we found the average fraction of contaminated reads across samples is 4.34% with a median value of 0.45% (Table S1). Specifically, 68% and 95% of individuals exhibited contaminated fractions below 5% and 15%, respectively (Fig. S12).

### Constructing the urinary metagenomic taxonomic and functional profile

Taxonomy assignment was performed using MetaPhlAn4 (59) version 4.0.6 with default settings. Compared to its predecessor, MetaPhlAn 3, MetaPhlAn 4 identifies a broader and more detailed range of urinary microbiota. Utilizing this tool, we obtained a raw microbial taxonomic data set composed of a total of 1,063 taxa (11 phyla, 32 classes, 60 orders, 115 families, 248 genera, and 597 species). The gene functional profiling was performed using HUMAnN v3.8 (60) with databases downloaded from <https://github.com/biobakery/humann/tree/v3.8>. Then, we identified a raw microbial pathway data set composed of 616 metabolic pathways. At last, we mainly used 208 taxa and 158 pathways with a relative abundance higher than 0.0001 and present in more than 10% of individuals for subsequent analysis.

### Source tracking analysis of urinary microbiome

We used R package “FEAST” to conduct source tracking analysis on 75 samples with microbiome sequencing data of all four body sites, by utilizing gut, saliva, and vaginal

microbiota as sources, with urinary microbiota serving as the sink (both in species level). Results were obtained for the percentage contribution of each body site.

### Calculating $\alpha$ - and $\beta$ -diversity, PCoA, and urotypes

The microbial diversity metrics, including  $\alpha$ -diversity (Shannon and Simpson indices) and  $\beta$ -diversity (Bray–Curtis dissimilarities), were calculated based on species-level abundance data using the “diversity” and “vegdist” functions within the R package “vegan,” respectively. Subsequently, principal coordinate analysis (PCoA) was conducted on the computed  $\beta$ -diversity dissimilarities utilizing the “capscale” function from the “vegan” package.

For sample clustering based on relative species abundances, the Jensen-Shannon divergence (JSD) distance metric and the PAM clustering algorithm (27) were employed. The optimal number of clusters was determined using the Calinski-Harabasz (CH) index (61) (Fig. S13). Following this, the LEfSe (29), a nonparametric cross-sample test utilizing LDA for test statistic computation, supported by traditional univariate tests for feature selection, was utilized. The microbial taxa and pathways were ranked based on their LDA values in descending order, and the predominant bacteria within each urotype were identified as the species with the highest LDA value.

### Host factors explained the variance of the urinary microbiome

We investigated the correlations between urinary microbiome compositions and 420 host variables encompassing anthropometric measurements, blood, and urine parameters, among others. To identify variables significantly linked to  $\beta$ -diversity, we employed Bray–Curtis distance-based redundancy analysis (dbRDA) and determined the variance explained by these factors using the “capscale” function in the vegan package. Significance of each response variable was verified through analysis of variance (ANOVA) for the dbRDA using the “anova.cca()” function in the vegan package. Only variables showing significant associations (Benjamini-Hochberg false discovery rate [FDR] < 0.05) with  $\beta$ -diversity in univariable models were included in the multivariable analysis. Due to some factors that were highly correlated, we calculated the Spearman correlations among host factors and removed those collineated factors (Spearman’s  $r > 0.6$  ||  $r < -0.6$ ). The remaining independent and significant factors were categorized into distinct groups including urinary parameters, anthropometric measures, health-related indicators, lifestyles, and dietary elements. To assess the combined explanatory power of these variable groups on urinary microbiome compositions, we performed a variation partitioning analysis using the vegan package. The “adj.r.squared” value was calculated using the RsquareAdj option to quantify the proportion of variance explained by the different categories.

We conducted a Pearson correlation analysis on the 158 metabolic pathways and 420 host factors using the cor.test() function in R (Table S7).

### High-depth whole-genome sequencing for the blood samples

Among the 1,579 individuals, 687 with blood samples were sequenced to a mean of 30 $\times$  for the whole genome. For blood samples, buffy coat was isolated, and DNA was extracted using HiPure Blood DNA Mini Kit (Magen, catalog no. D3111) according to the manufacturer’s protocol. The library construction and sequencing methods were consistent with those described in a previous article (19).

For genomic analysis, whole-genome reads were aligned to the GRCh38/hg38 human genome using BWA (62) (v0.7.15), with low-quality reads filtered out. Alignments were processed into BAM files using Samtools (63) (v0.1.18), and PCR duplicates were marked with Picardtools (v1.62). The Genome Analysis Toolkit (GATK [64], v3.8) facilitated SNP and INDEL identification and recalibration, referencing dbSNP (v150). Variant discovery and processing were performed with GATK’s HaplotypeCaller and subsequent tools. The final data set, comprising 7.2 million SNPs/INDELS (MAF  $\geq$  1%), was left for M-GWAS analyses.

For variant inclusion, we set stringent thresholds: (i) a minimum average depth of over 8x; (ii) a Hardy-Weinberg equilibrium (HWE)  $P$  value  $> 10^{-5}$ ; and (iii) a genotype calling rate  $> 98\%$ . Regarding sample criteria, we required the following: (i) a mean sequencing depth  $> 20x$ ; (ii) a variant call rate  $> 98\%$ ; (iii) the absence of population stratification, confirmed through principal component analysis (PCA) conducted using PLINK (65) (version 1.07); and (iv) the exclusion of related individuals, determined by calculating pairwise identity by descent (IBD, with a Pi-hat threshold of 0.1875) in PLINK. Following these rigorous quality control measures, 687 individuals were left for subsequent analysis.

### Sex-stratified analysis and predictive analysis using urinary microbiota

We used LEfSe analysis to discover sex-differentiated microbiota by applying an LDA threshold of 2.0 and a significance cutoff less than 0.05. Furthermore, we employed the random forest algorithm to explore the potential of 47 sex-differentiated microbiota (18 genus and 29 species) for gender prediction. This approach was chosen due to its efficacy in handling complex biological data and its robustness in feature selection, making it particularly suitable for high-dimensional data sets like microbiome compositions. We first divided our data set into 70% training and 30% testing sets, ensuring a balanced representation of male and female samples. The random forest model was then trained on the training set with gender as the response variable and the average relative abundance of different urinary microbes as predictor variables. We used 500 trees in our model to ensure adequate learning and generalization capability. The model's performance was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), which provides a comprehensive measure of model accuracy across different classification thresholds. Furthermore, we ranked urinary microbiome according to Gini importance values.

### Association analysis for microbial taxa

Considering the power of GWAS tests, we refined our approach to focus on microbial taxa exhibiting occurrence rates exceeding 10% (present in at least 157 individuals) and an average relative abundance above  $1 \times 10^{-4}$ . After filtering, the represented genera of these microbial taxa covered 99.73% of the whole community in the cohort. Subsequently, a total of 106 microbial genera and species were included for association analyses.

We examined the relationships between host genetics and the urine microbiome using two different statistical models: a linear model based on relative abundance and a logistic model based on the presence/absence (P/A) of microbial features. For those taxa present in over 50% individuals, we used a linear model and transformed the relative abundance data using a natural logarithm. Subsequently, we calculated the residuals by employing the "lm" function in R:  $\log_{10}(\text{microbe abundance}) - \text{age} + \text{sex} + \text{BMI} + \text{sequencing read counts} + \text{top ten PCs}$ . These residuals were then extracted using the residuals() function from the R stats package and were applied in a univariate linear model for association analysis with genotypes. For the microbial taxa observed in more than 10% but less than 90% of individuals, we converted the data into presence/absence patterns to avoid issues with zero inflation. By dichotomizing these data, we could treat the bacterial abundance as a binary trait for logistic regression analysis. This analysis was adjusted for the same set of covariates as mentioned earlier. This dual approach allowed us to comprehensively assess the influence of host genetics on different aspects of the urine microbiome.

We searched the oral microbiome-related SNPs in the summary statistics data from this cohort as previously reported to examine their associations with host traits.

Additionally, we conducted GWAS association analyses similar to the steps described above for  $\alpha$ - and  $\beta$ -diversity and urinary clusters; however, no genome-wide significant signals were observed (Tables S13 and S14). We also performed gene–environment interaction analyses on the top two significant loci, chr8:109442613 and chr2:191850055.

These analyses, focusing on interactions with gender and serum testosterone levels, were executed using the --interaction parameter in PLINK.

## Two-sample MR analysis

To explore the causal relationships between the urinary microbiome and various diseases, we accessed the summary statistics data of 42 diseases from Biobank Japan (40). The two-sample bidirectional MR analysis was performed by applying the generalized summary-data-based Mendelian randomization (GSMR) method (66). Genetic variants with  $P < 1 \times 10^{-5}$  and LD  $r^2 < 0.1$  were also selected as instrumental variables for phenotypes in the Japan Biobank study.

## ACKNOWLEDGMENTS

We sincerely thank the support provided by the China National GeneBank. We thank all the volunteers for their time and for self-collecting the oral samples using our kit.

This work was supported by the National Natural Science Foundation of China (No. 32200548).

X.L. and T.Z. conceived and organized this study. J.W. initiated the overall health project. X.X., X.J., L.X., Y.H., and Y.Z. contributed to organization of the cohort, the sample collection, and questionnaire collection. H.L. led the DNA extraction and sequencing. L.Z. and Z.Z. processed the metagenome data. X.L. and X.T. processed the whole-genome data. L.Z. and X.L. performed the metagenome-genome-wide association analyses. L.Z., Z.Z., and X.L. wrote the raw manuscript. All authors contributed to data and texts in the manuscript.

## AUTHOR AFFILIATIONS

<sup>1</sup>BGI Research, Wuhan, China

<sup>2</sup>BGI Research, Shenzhen, China

<sup>3</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

<sup>4</sup>China National Genebank, BGI-Shenzhen, Shenzhen, China

<sup>5</sup>James D. Watson Institute of Genome Sciences, Hangzhou, China

<sup>6</sup>Shenzhen Engineering Laboratory of Detection and Intervention of Human Intestinal Microbiome, BGI-Shenzhen, Shenzhen, China

<sup>7</sup>Institute of Precision Medicine–Greater Bay Area (Guangzhou), Fudan University, Guangzhou, China

<sup>8</sup>Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, BGI Research, Shenzhen, China

## AUTHOR ORCIDs

Leying Zou  <http://orcid.org/0000-0002-7243-8410>

Junhong Chen  <http://orcid.org/0000-0002-8313-8317>

Liang Xiao  <http://orcid.org/0000-0003-0836-4397>

Huijue Jia  <http://orcid.org/0000-0002-3592-126X>

Tao Zhang  <http://orcid.org/0000-0003-2765-2802>

Xiaomin Liu  <http://orcid.org/0000-0003-2155-3627>

## AUTHOR CONTRIBUTIONS

Leying Zou, Data curation, Methodology, Project administration, Software, Visualization, Writing – original draft | Zhe Zhang, Data curation, Writing – review and editing | Xiaomin Liu, Data curation, Supervision, Writing – review and editing, Writing – original draft.

## DATA AVAILABILITY

The metagenomic sequencing data after removing host reads have been deposited to the CNGB Nucleotide Sequence Archive (CNSA: <https://db.cngb.org/cnsa>; accession

number [CNP0003862](https://db.cncb.org/search/project/CNP0003862)). The summary statistics that support the findings of this study, including the associations between host genetics and urinary microbiome, are available from <https://db.cncb.org/search/project/CNP0005626/>. According to the Human Genetic Resources Administration of China regulation and the institutional review board of BGI-Shenzhen related to protecting individual privacy, sequencing data are controlled access and are available via application on request.

## ETHICS APPROVAL

All recruitment and study procedures involving human subjects were approved by the Institutional Review Boards (IRBs) at BGI-Shenzhen (IRB number: 19121). All participants provided written informed consent at enrollment.

## ADDITIONAL FILES

The following material is available online.

### Supplemental Material

**Supplemental Figures (mBio02773-24-s0001.docx).** Fig. S1 to S13.

**Supplemental Tables (mBio02773-24-s0002.xlsx).** Tables S1 to S14.

## REFERENCES

- Wolfe AJ, Brubaker L. 2019. Urobiome updates: advances in urinary microbiome research. *Nat Rev Urol* 16:73–74. <https://doi.org/10.1038/s41585-018-0127-5>
- Whiteside SA, Razvi H, Dave S, Reid G, Burton JP. 2015. The microbiome of the urinary tract—a role beyond infection. *Nat Rev Urol* 12:81–90. <https://doi.org/10.1038/nrurol.2014.361>
- Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *MBio* 5:e01283–01214. <https://doi.org/10.1128/mBio.01283-14>
- Ishihara T, Watanabe N, Inoue S, Aoki H, Tsuji T, Yamamoto B, Yanagi H, Oki M, Kryukov K, Nakagawa S, Inokuchi S, Ozawa H, Imanishi T. 2020. Usefulness of next-generation DNA sequencing for the diagnosis of urinary tract infection. *DD&T* 14:42–49. <https://doi.org/10.5582/ddt.2020.01000>
- Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, Brubaker L, Wolfe AJ, Mueller ER, Schreckenberger PC. 2016. The clinical urine culture: enhanced techniques improve detection of clinically relevant microorganisms. *J Clin Microbiol* 54:1216–1222. <https://doi.org/10.1128/JCM.00044-16>
- Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD. 2018. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat Commun* 9:1557. <https://doi.org/10.1038/s41467-018-03968-5>
- Komesu YM, Richter HE, Carper B, Dinwiddie DL, Lukacz ES, Siddiqui NY, Sung VW, Zyczynski HM, Ridgeway B, Rogers RG, Arya LA, Mazloomdoost D, Gantz MG, Pelvic Floor Disorders Network. 2018. The urinary microbiome in women with mixed urinary incontinence compared to similarly aged controls. *Int Urogynecol J* 29:1785–1795. <https://doi.org/10.1007/s00192-018-3683-6>
- Govender Y, Gabriel I, Minassian V, Fichorova R. 2019. The current evidence on the association between the urinary microbiome and urinary incontinence in women. *Front Cell Infect Microbiol* 9:133. <https://doi.org/10.3389/fcimb.2019.00133>
- Moustafa A, Li W, Singh H, Moncera KJ, Torralba MG, Yu Y, Manuel O, Biggs W, Venter JC, Nelson KE, Pieper R, Telenti A. 2018. Microbial metagenome of urinary tract infection. *Sci Rep* 8:4333. <https://doi.org/10.1038/s41598-018-22660-8>
- Vaughan MH, Mao J, Karstens LA, Ma L, Amundsen CL, Schmader KE, Siddiqui NY. 2021. The urinary microbiome in postmenopausal women with recurrent urinary tract infections. *J Urol* 206:1222–1231. <https://doi.org/10.1097/JU.0000000000001940>
- Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol* 13:269–284. <https://doi.org/10.1038/nrmicro3432>
- Thomas-White KJ, Gao X, Lin H, Fok CS, Ghanayem K, Mueller ER, Dong Q, Brubaker L, Wolfe AJ. 2018. Urinary microbes and postoperative urinary tract infection risk in urogynecologic surgical patients. *Int Urogynecol J* 29:1797–1805. <https://doi.org/10.1007/s00192-018-3767-3>
- Wu P, Zhang G, Zhao J, Chen J, Chen Y, Huang W, Zhong J, Zeng J. 2018. Corrigendum: profiling the urinary microbiota in male patients with bladder cancer in China. *Front Cell Infect Microbiol* 8:429. <https://doi.org/10.3389/fcimb.2018.00429>
- Bučević Popović V, Šitum M, Chow C-ET, Chan LS, Roje B, Terzić J. 2018. The urinary microbiome associated with bladder cancer. *Sci Rep* 8:12157. <https://doi.org/10.1038/s41598-018-29054-w>
- van Puffelen JH, Keating ST, Oosterwijk E, van der Heijden AG, Netea MG, Joosten LAB, Vermeulen SH. 2020. Trained immunity as a molecular mechanism for BCG immunotherapy in bladder cancer. *Nat Rev Urol* 17:513–525. <https://doi.org/10.1038/s41585-020-0346-4>
- Qin J, Shi X, Xu J, Yuan S, Zheng B, Zhang E, Huang G, Li G, Jiang G, Gao S, Tian C, Guo R, Fu Z, Huang Q, Yang R, Zhang W, Li S, Wu S. 2021. Characterization of the genitourinary microbiome of 1,165 middle-aged and elderly healthy individuals. *Front Microbiol* 12:673969. <https://doi.org/10.3389/fmcb.2021.673969>
- Ammitzbøll N, Bau BPJ, Bundgaard-Nielsen C, Villadsen AB, Jensen A-M, Leutscher PDC, Glavind K, Hagstrøm S, Arenholt LTS, Sørensen S. 2021. Pre- and postmenopausal women have different core urinary microbiota. *Sci Rep* 11:2212. <https://doi.org/10.1038/s41598-021-81790-8>
- Adebayo AS, Ackermann G, Bowyer RCE, Wells PM, Humphreys G, Knight R, Spector TD, Steves CJ. 2020. The urinary tract microbiome in older women exhibits host genetic and environmental influences. *Cell Host Microbe* 28:298–305. <https://doi.org/10.1016/j.chom.2020.06.022>
- Liu X, Tang S, Zhong H, Tong X, Jie Z, Ding Q, Wang D, Guo R, Xiao L, Xu X, Yang H, Wang J, Zong Y, Liu W, Liu X, Zhang Y, Brix S, Kristiansen K, Hou Y, Jia H, Zhang T. 2021. A genome-wide association study for gut metagenome in Chinese adults illuminates complex diseases. *Cell Discov* 7. <https://doi.org/10.1038/s41421-020-00239-w>
- Liu X, Tong X, Zhu J, Tian L, Jie Z, Zou Y, Lin X, Liang H, Li W, Ju Y, Qin Y, Zou L, Lu H, Zhu S, Jin X, Xu X, Yang H, Wang J, Zong Y, Liu W, Hou Y, Jia H, Zhang T. 2021. Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. *Cell Discov* 7:117. <https://doi.org/10.1038/s41421-021-00356-0>

21. Liu X, Tong X, Zou Y, Lin X, Zhao H, Tian L, Jie Z, Wang Q, Zhang Z, Lu H, Xiao L, Qiu X, Zi J, Wang R, Xu X, Yang H, Wang J, Zong Y, Liu W, Hou Y, Zhu S, Jia H, Zhang T. 2022. Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nat Genet* 54:52–61. <https://doi.org/10.1038/s41588-021-00968-y>
22. Jie Z, Liang S, Ding Q, Li F, Tang S, Wang D, Lin Y, Chen P, Cai K, Qiu X, et al. 2021. A transomic cohort as A reference point for promoting A healthy human gut microbiome. *Med Microecol* 8:100039. <https://doi.org/10.1016/j.medmic.2021.100039>
23. Zhu J, Tian L, Chen P, Han M, Song L, Tong X, Sun X, Yang F, Lin Z, Liu X, Liu C, Wang X, Lin Y, Cai K, Hou Y, Xu X, Yang H, Wang J, Kristiansen K, Xiao L, Zhang T, Jia H, Jie Z. 2022. Over 50,000 metagenomically assembled draft genomes for the human oral microbiome reveal new taxa. *Genomics Proteomics Bioinformatics* 20:246–259. <https://doi.org/10.1016/j.gpb.2021.05.001>
24. Chen C, Hao L, Zhang Z, Tian L, Zhang X, Zhu J, Jie Z, Tong X, Xiao L, Zhang T, Jin X, Xu X, Yang H, Wang J, Kristiansen K, Jia H. 2021. Cervicovaginal microbiome dynamics after taking oral probiotics. *J Genet Genom* 48:716–726. <https://doi.org/10.1016/j.jgg.2021.03.019>
25. Huttenhower C. 2012. Structure, function and diversity of the healthy human microbiome. *Nature New Biol* 486:207–214. <https://doi.org/10.1038/nature11234>
26. Shenhav L, Thompson M, Joseph TA, Briscoe L, Furman O, Bogumil D, Mizrahi I, Pe'er I, Halperin E. 2019. FEAST: fast expectation-maximization for microbial source tracking. *Nat Methods* 16:627–632. <https://doi.org/10.1038/s41592-019-0431-x>
27. Kaufman L, Rousseeuw PJ. 1990. Partitioning around medoids (Program PAM), p 68–125. In *Finding groups in data*. <https://doi.org/https://doi.org/10.1002/9780470316801.ch2>.
28. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, et al. 2011. Enterotypes of the human gut microbiome. *Nature New Biol* 473:174–180. <https://doi.org/10.1038/nature09944>
29. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol* 12:R60. <https://doi.org/10.1186/gb-2011-12-6-r60>
30. Norenhag J, Edfeldt G, Stålberg K, Garcia F, Hugerth LW, Engstrand L, Fransson E, Du J, Schuppe-Koistinen I, Olovsson M. 2024. Compositional and functional differences of the vaginal microbiota of women with and without cervical dysplasia. *Sci Rep* 14:11183. <https://doi.org/10.1038/s41598-024-61942-2>
31. Mosca L, Minopoli M, Pagano M, Vitiello F, Carriero MV, Cacciapuoti G, Porcelli M. 2020. Effects of S-adenosyl-L-methionine on the invasion and migration of head and neck squamous cancer cells and analysis of the underlying mechanisms. *Int J Oncol* 56:1212–1224. <https://doi.org/10.3892/ijo.2020.5011>
32. Zhang X, Zhong H, Li Y, Shi Z, Ren H, Zhang Z, Zhou X, Tang S, Han X, Lin Y, Yang F, Wang D, Fang C, Fu Z, Wang L, Zhu S, Hou Y, Xu X, Yang H, Wang J, Kristiansen K, Li J, Ji L. 2021. Sex- and age-related trajectories of the adult human gut microbiota shared across populations of different ethnicities. *Nat Aging* 1:87–100. <https://doi.org/10.1038/s43587-020-0014-2>
33. Graham ME, Herbert WG, Song SD, Raman HN, Zhu JE, Gonzalez PE, Walther-António MRS, Tetel MJ. 2021. Gut and vaginal microbiomes on steroids: implications for women's health. *Trends Endocrinol Metab* 32:554–565. <https://doi.org/10.1016/j.tem.2021.04.014>
34. Henningsson SSG, Rosengren E. 1972. Alterations of histamine metabolism after injections of sex hormones in mice. *British J Pharmacology* 44:517–526. <https://doi.org/10.1111/j.1476-5381.1972.tb07288.x>
35. Giltay EJ, Bunck MC, Gooren LJ, Zitman FG, Diamant M, Teerlink T. 2008. Effects of sex steroids on the neurotransmitter-specific aromatic amino acids phenylalanine, tyrosine, and tryptophan in transsexual subjects. *Neuroendocrinology* 88:103–110. <https://doi.org/10.1159/000135710>
36. Yang Y, Pang Q, Hua M, Huangfu Z, Yan R, Liu W, Zhang W, Shi X, Xu Y, Shi J. 2023. Excavation of diagnostic biomarkers and construction of prognostic model for clear cell renal cell carcinoma based on urine proteomics. *Front Oncol* 13. <https://doi.org/10.3389/fonc.2023.1170567>
37. Onuchic LF, Furu L, Nagasawa Y, Hou X, Eggermann T, Ren Z, Bergmann C, Senderek J, Esquivel E, Zeltner R, Rudnik-Schöneborn S, Mrug M, Sweeney W, Avner ED, Zerres K, Guay-Woodford LM, Somlo S, Germino GG. 2002. PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix 1 repeats. *Am J Hum Genet* 70:1305–1317. <https://doi.org/10.1086/340448>
38. Hansen CG, Bright NA, Howard G, Nichols BJ. 2009. SDPR induces membrane curvature and functions in the formation of caveolae. *Nat Cell Biol* 11:807–814. <https://doi.org/10.1038/ncb1887>
39. Costa VL, Henrique R, Danielsen SA, Duarte-Pereira S, Eknaes M, Skotheim RI, Rodrigues A, Magalhães JS, Oliveira J, Lothe RA, Teixeira MR, Jerônimo C, Lind GE. 2010. Three epigenetic biomarkers, GDF15, TMEFF2, and VIM, accurately predict bladder cancer from DNA-based analyses of urine samples. *Clin Cancer Res* 16:5842–5851. <https://doi.org/10.1158/1078-0432.CCR-10-1312>
40. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Kubo M, Okada Y, Kamatani Y. 2018. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 50:390–400. <https://doi.org/10.1038/s41588-018-0047-6>
41. Komatsu YM, Dinwiddie DL, Richter HE, Lukacz ES, Sung VW, Siddiqui NY, Zyczynski HM, Ridgeley B, Rogers RG, Arya LA, Mazloomdoost D, Levy J, Carper B, Gantz MG, Unice Kennedy Shriver National Institute of Child Health and Human Development Pelvic Floor Disorders Network. 2020. Defining the relationship between vaginal and urinary microbiomes. *Am J Obstet Gynecol* 222:154. <https://doi.org/10.1016/j.ajog.2019.08.011>
42. Buhmann MT, Abt D, Nolte O, Neu TR, Strempel S, Albrich WC, Betschart P, Zumstein V, Neels A, Maniura-Weber K, Ren Q. 2019. Encrustations on ureteral stents from patients without urinary tract infection reveal distinct urotypes and a low bacterial load. *Microbiome* 7:60. <https://doi.org/10.1186/s40168-019-0674-x>
43. Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. 2011. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. *BMC Microbiol* 11:244. <https://doi.org/10.1186/1471-2180-11-244>
44. Fouts DE, Pieper R, Szpakowski S, Pohl H, Knoblauch S, Suh M-J, Huang S-T, Ljungberg I, Sprague BM, Lucas SK, Torralba M, Nelson KE, Groah SL. 2012. Integrated next-generation sequencing of 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic bacteriuria in neuropathic bladder associated with spinal cord injury. *J Transl Med* 10:174. <https://doi.org/10.1186/1479-5876-10-174>
45. Lewis DA, Brown R, Williams J, White P, Jacobson SK, Marchesi JR, Drake MJ. 2013. The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. *Front Cell Infect Microbiol* 3:41. <https://doi.org/10.3389/fcimb.2013.00041>
46. Gottschick C, Deng Z-L, Vital M, Masur C, Abels C, Pieper DH, Wagner-Döbler I. 2017. The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. *Microbiome* 5:99. <https://doi.org/10.1186/s40168-017-0305-3>
47. Foxman B. 2010. The epidemiology of urinary tract infection. *Nat Rev Urol* 7:653–660. <https://doi.org/10.1038/nrurol.2010.190>
48. Stamm WE. 2002. Scientific and clinical challenges in the management of urinary tract infections. *Am J Med* 113 Suppl 1A:1S–4S. [https://doi.org/10.1016/s0002-9343\(02\)01053-7](https://doi.org/10.1016/s0002-9343(02)01053-7)
49. Odoki M, Almustapha Aliero A, Tibyangye J, Nyabayayo Maniga J, Wampande E, Drago Kato C, Agwu E, Bazira J. 2019. Prevalence of bacterial urinary tract infections and associated factors among patients attending hospitals in Bushenyi District, Uganda. *Int J Microbiol* 2019:4246780. <https://doi.org/10.1155/2019/4246780>
50. Björnsson TD. 1979. Use of serum creatinine concentrations to determine renal function. *Clin Pharmacokinet* 4:200–222. <https://doi.org/10.2165/00003088-197904030-00003>
51. Moitinho-Silva L, Degenhardt F, Rodriguez E, Emmert H, Juzenas S, Möbus L, Uellendahl-Werth F, Sander N, Baurecht H, Tittmann L, Lieb W, Gieger C, Peters A, Ellinghaus D, Bang C, Franke A, Weidinger S, Rühlemann MC. 2022. Host genetic factors related to innate immunity, environmental sensing and cellular functions are associated with human skin microbiota. *Nat Commun* 13:6204. <https://doi.org/10.1038/s41467-022-33906-5>
52. Cartwright R, Franklin L, Tikkinen KAO, Kalliala I, Miotla P, Rechberger T, Offiah I, McMahon S, O'Reilly B, Lince S, et al. 2021. Genome-wide

- association study identifies two novel loci associated with female stress and urgency urinary incontinence. *J Urol* 206:679–687. <https://doi.org/10.1097/JU.0000000000001822>
53. Qin J, Li Y, Cai Z, Li 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 New Biol* 490:55–60. <https://doi.org/10.1038/nature11450>
54. Lilan H, Lilan H. 2020. DNA extraction for human microbe samples. *protoc* io
55. Han M, Hao L, Lin Y, Li F, Wang J, Yang H, Xiao L, Kristiansen K, Jia H, Li J. 2018. A novel affordable reagent for room temperature storage and transport of fecal samples for metagenomic analyses. *Microbiome* 6:43. <https://doi.org/10.1186/s40168-018-0429-0>
56. Ackerman AL, Anger JT, Khalique MU, Ackerman JE, Tang J, Kim J, Underhill DM, Freeman MR, NIH Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP). 2019. Optimization of DNA extraction from human urinary samples for mycobiome community profiling. *PLoS One* 14:e0210306. <https://doi.org/10.1371/journal.pone.0210306>
57. Yang F, Sun J, Luo H, Ren H, Zhou H, Lin Y, Han M, Chen B, Liao H, Brix S, Li J, Yang H, Kristiansen K, Zhong H. 2020. Assessment of fecal DNA extraction protocols for metagenomic studies. *Gigascience* 9:12. <https://doi.org/10.1093/gigascience/giaa071>
58. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6:226. <https://doi.org/10.1186/s40168-018-0605-2>
59. Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas AM, et al. 2023. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol* 41:1633–1644. <https://doi.org/10.1038/s41587-023-01688-w>
60. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Maiyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 10:e65088. <https://doi.org/10.7554/elife.65088>
61. Calinski T, Harabasz J. 1974. A dendrite method for cluster analysis. *Comm in Stats - Theory & Methods* 3:1–27. <https://doi.org/10.1080/03610927408827101>
62. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
63. Li Heng, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
64. McKenna N, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
65. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
66. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, Robinson MR, McGrath JJ, Visscher PM, Wray NR, Yang J. 2018. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 9:224. <https://doi.org/10.1038/s41467-017-02317-2>