Differential Microbial Signature Associated With Benign Prostatic Hyperplasia and Prostate Cancer

Purandar Sarkar
Presidency University
Samaresh Malik
Presidency University
Anwesha Banerjee
Presidency University
Chhanda Datta
Institute of Post Graduate Medical Education and Research
Dilip Kumar Pal
Institute of Post Graduate Medical Education and Research
Amlan Ghosh
Presidency University
Abhik Saha (abhik.dbs@presiuniv.ac.in)
Presidency University

Research Article

Keywords:

Posted Date: February 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1318477/v1

Abstract

Apart from other risk factors, chronic inflammation is also associated with the onset of Prostate Cancer (PCa), wherein pathogen infection and tissue microbiome dysbiosis are known to play a major role for both inflammatory response and cancer development. However, except for a few studies, the link between microbes and PCa remained poorly understood. To explore the potential microbiome signature associated with PCa in Indian patients, we investigated differential compositions of commensal bacteria among patients with benign prostatic hyperplasia (BPH) and PCa using 16S rRNA amplicon sequencing followed by qPCR analyses using two distinct primer sets. Using two independent cohorts, we show that *Prevotella copri*, *Cupriavidus campinensis* and *Propionibacterium acnes* represent three most abundant bacteria in diseased prostate lesions. LEfSe analyses identified that while *Cupriavidus taiwanensis and Methylobacterium organophilum* are distinctly elevated in PCa samples, *Kocuria palustris* and *Cellvibrio mixtus* are significantly enriched in BPH samples. Furthermore, we identify that a number of human tumor viruses including Epstein-Barr virus (EBV), hepatitis B virus (HBV) along with two high risk human papillomaviruses - HPV-16 and HPV-18 are significantly associated with the PCa development and strongly correlated with PCa bacterial signature. The study may thus offer to develop a framework for exploiting this microbial signature for early diagnosis and prognosis of PCa development.

Introduction

Prostate cancer (PCa) is the second most common cancer and the sixth leading cause of cancer associated deaths among men, leading to a great public health concern worldwide[1]. Recently, due to the increase of life expectancy, adoption of newer lifestyles and changes in food habit, the incidence rate of PCa is gradually increasing in India[2]. Mounting evidence suggests that chronic inflammation plays a vital role for the initiation and progression of PCa[3,4]. Prostate gland harbors a relatively higher number of lymphocytic population and other immune cells for protecting the male reproductive system from invading pathogens through bladder and urethra[3,5]. Particularly in case of elderly male population, a significant increase of the immune cell population in the prostate leads to chronic and acute inflammatory condition, and thereby increasing the likelihood of PCa development. Prostate specific antigen (PSA) testing followed by trans-urethral resection of the prostate has long been recommended for diagnosis of prostate inflammation, while prognosis of PCa is predicted by Gleason grading, a low-power microscopic evaluation subjected to inter-observer variations and lacks precision[6,7]. However, through this method, detection of PCa even at an early stage is not always accompanied by accurate determination of morbidity risk and therefore mortality might result due to over-treatment in some cases and under-treatment in others. Thus, a great attention has been made thereafter by the researchers to explore better molecular approaches for the early diagnosis and prognosis of PCa.

The advent of several Next-Generation sequencing technologies has led to a new era of unbiased identification and characterization of complex microbial communities as well as its association with several pathological conditions such as chronic inflammation and its possible role in cancer

development[3,8]. Pathogen including oncogenic viruses and bacterial infections are thought to play critical role in prostatic inflammation[9]. However, hitherto no single microorganism has appeared as a direct contributor of chronic inflammation in PCa patients. Nevertheless, a growing body of evidence suggests that a number of bacterial species including *E. coli* and other species of Enterobacteriaceae can promote prostatic inflammation[3,10]. In addition, sexually transmitted infections with *Chlamydia trachomatis*, *Neisseria gonorrhoeae* and *Trichomonas vaginalis* are also shown to increase prostatic inflammation, which in turn elevate serum PSA levels[11,12]. The pro-inflammatory *Propionibacterium* species, in particular *P. acnes* associated with human skin, has been well studied in connection to PCa development, and has also been found to induce prostatic inflammation in animal models[13–16]. Beside tissue microbiome dysbiosis, tumor virus infections are also shown to play a major role in cancer development, accounting to approximately 20% of all human cancers[17]. Studies have shown that a number of viruses including multiple subtypes of human papillomaviruses (HPVs), polyomaviruses such as John Cunningham virus (JCV), BK virus (BKV), and simian virus 40 (SV40) as well as herpesviruses like human cytomegalovirus (HCMV) and Epstein-Barr virus (EBV) are associated with PCa specimens[18–22].

Recently a number of Next-Generation sequencing as well as hybridization based microarray studies identified potential microbiome signature associated with PCa development[19,23–25]. Despite these efforts in identifying the composition of the microbiome in prostate tissue and its implication of disease progression, the pathogenic microbial composition of PCa may vary among the population of different ethnicity across the globe. Given the increasing incidence of PCa patients in India[26], a detailed and comprehensive analysis of the microbial ecosystem coupled with the PCa development is of utmost importance. Herein, we aim to identify specific microbial signature associated with the pathologic prostate tissue specimens collected from patients in the Eastern region of India.

Results

Sample characteristics and sequencing data summary of diseased prostate samples: To understand the changes of microbial compositions associated with prostate cancer (PCa) development, we prospectively collected a total 46 tissue biopsy samples from 13 benign prostatic hyperplasia (BPH) and 33 PCa patients (‘Discovery Cohort’, Table S1). The total DNA of each specimen from the ‘Discovery Cohort’ was extracted and subjected to 16S rRNA amplicon based sequencing on an Ion GeneStudio S5 System targeting the hypervariable regions of 16S rRNA gene using two distinct primer sets as per manufacturer’s instructions. While Set I primer was used for amplification of hypervariable regions including V2, V4 and V8, set II primer was used for hypervariable regions including V3, V6-7 and V9. Additionally, as described later, in order to validate the sequencing data, we also collected another 31 tissue biopsy samples from 16 BPH and 15 PCa patients (‘Validation Cohort’, Table S1).

A total of 35,320,332 raw reads were generated after sequencing of the 46 samples from the ‘Discovery Cohort’. The number of raw sequence reads varied by approximately 10 fold across samples. After quality trimming and chimera checking, 30,641,026 high quality reads were mapped to two comprehensive 16S

rRNA reference databases – Greengenes[27] (v13.5) and the Thermo Fisher Scientific in-house MicroSeq 500[28] (v2013.1). 6,301,956 reads were excluded from the analyses due to low copy number reads (less than 10 copy numbers) and 41,312 reads were found to be un-mapped in both databases (Table S2). 24,297,758 mapped reads subsequently identified Operational Taxonomic Units (OTUs) at the level of family (95% and above similarity index), genus (97% and above similarity index) and species (99% and above similarity index). Overall, a total of 6 phylum, 127 genera and 291 bacterial species were identified from all 46 samples in the discovery cohort.

Overall bacterial abundance and diversity among BPH and PCa samples: FASTQ files from 16S rRNA gene amplicon sequencing data of all 46 samples were analyzed using a web-based online tool - MicrobiomeAnalyst for statistical analysis[29]. Rarefaction curves of species richness against sequences per sample were plotted for BPH and PCa samples to determine the efficiency of the sequencing process (Fig. 1A-B). Most of the samples, though not completely, reached a saturated plateau phase, indicating that the depth of sequencing was sufficient for the diversity analysis (Fig. 1A-B). Both observed species (p = 0.015) and Chao1 index (p = 0.034) showed that species richness was significantly decreased in PCa samples as compared to BPH samples (Figs. 1C and 1D, respectively). In addition, the diversity estimators both Shannon index (p = 0.158) and Simpson index (p = 0.411) indicated a trend of depletion in relative diversity of species composition in PCa samples as compared to BPH samples, although the data was not statistically significant (Figs. 1E and 1F, respectively).

To assess the diversity among two groups, we evaluated weighted UniFrac distance matrix [30] from the OTU abundance through utilizing Permutational Multivariate Analysis of Variance (PERMANOVA) algorithm [31] and subsequently applied in Principal Component Analysis (PCoA) [32] (Fig. 1G). Given that PCoA analyses revealed no significant difference (p = 0.082) in the bacterial compositions between BPH and PCa samples (Fig. 1G), a 'Random Forest' algorithm was applied to further confirm the difference in bacterial community among the BPH and PCa biopsy samples (Fig. 1H). The decision trees extracted from the random forest classification identified distinct bacterial composition in 2 BPH samples and 11 BPH samples exhibited overlapping species with PCa samples (class error: 0.850; Fig. 1H). In contrast, all 33 PCa samples demonstrated unique bacterial compositions (Fig. 1H).

Taxonomic characterization of bacterial compositions among BPH and PCa samples: The bacterial communities associated with the diseased prostate lesions were further analyzed at different taxonomic levels. Six phyla including *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Fusobacteria* and *Deinococcus-Thermus* collectively comprised of the entire sequences in both groups (Fig. 2A-B). In general, *Proteobacteria* was the most abundant phylum in the prostate microbial ecology, contributing ~40.6% in diseased prostate lesions (Fig. 2A-B). Overall, the results indicated that only *Actinobacteria* phylum was significantly depleted in PCa samples as compared to the BPH category (Fig. 2A-B). At the genus level, *Prevotella*, *Cupriavidus*, *Propionibacterium*, *Acinetobacter* and *Corynebacterium* represented the top five genera in diseased prostate lesion, comprising of 19.94%, 13.54%, 6.54%, 5.8% and 5.23% sequence coverage, respectively in BPH and 26.85%, 14.07%, 3.97%, 8.78% and 3.5% sequence coverage,

respectively in PCa samples (Figs. 2C-D and S1). Of all genera detected, both groups shared approximately half of the total genera identified i.e. 56/107 in BPH category and 56/118 in PCa lesions (Fig. S1).

To identify the differentially enriched genera within BPH and PCa samples, we employed Linear Discrimination Analysis (LDA) Effect Size (LEfSe) method [33] (Fig. 2E). LEfSe analysis of top 20 bacterial genera identified *Kocuria*, *Staphylococcus*, *Corynebacterium*, *Cellvibrio*, *Pseudomonas*, *Paracoccus*, *Brachybacterium*, *Pseudoxanthomonas*, *Anaerococcus*, *Stenotrophomonas*, *Microvirga*, *Empedobacter*, *Lysobacter*, *Brevibacterium*, *Comamonas*, *Serinicoccus*, *Rhodobacter*, *Chryseobacterium* and *Aeromicrobium* as enriched genera in BPH samples, whereas only *Bradyrhizobium* genus was found to be significantly elevated in PCa samples (Fig. 2E). Overall, the results indicated that enriched diversity of bacterial taxa within diseased prostate lesions was mostly contributed by BPH samples.

Comparable enrichment analysis at species level identified significant bacterial compositions between BPH and PCa samples: Detailed analyses of bacterial compositions at species level demonstrated *Prevotella copri*, *Cupriavidus campinensis*, *Propionibacterium acnes* and Paracoccus sp. covered almost 50% of species variety among diseased prostate lesions including both BPH and PCa samples (Figs. 3A-B). Next, as similar to the genus level, cladogram and LEfSe analyses were conducted to further uncover the differential species compositions between BPH and PCa tissue biopsy samples (Figs. 3C-D). The LDA scores demonstrated that among top 20 most significantly enriched species *Kocuria palustris, Cellvibrio mixtus, Pseudomonas stutzeri, Paracoccus sp, Staphylococcus hominis*, *Corynebacterium tuberculostearicum*, *Brachybacterium paraconglomeratum*, *Staphylococcus arlettae*, *Staphylococcus cohnii* and *Anaerococcus octavius* were notably elevated in the BPH samples, while *Cupriavidus taiwanensis*, *Methylobacterium organophilum*, *Brevundimonas vancanneytii*, *Neisseria flavescens*, *Acinetobacter junii*, *Bradyrhizobium cytisi*, *Cupriavidus basilensis*, *Caulobacter segnis*, *Leclercia adecarboxylata* and *Neisseria elongata* were significantly increased in the PCa samples (Fig. 3D).

Validation of species identified in 16S rRNA sequencing by quantitative real-time PCR analyses in BPH and PCa samples: Although 16S rRNA amplicon based sequencing can produce robust data regarding the presence and abundance of bacterial species, horizontal gene transfer, multiple copies of 16S rRNA and other limitations can generate error prone abundance data [8,34]. To confirm the 16S rRNA sequencing data, quantitative real-time PCR (qPCR) analyses were performed of top 2 bacteria in each category – BPH (*K. palustris* and *C. mixtus*) and PCa (*C. taiwanensis* and *Methylobacterium organophilum*), along with 3 most abundant bacteria (*P. copri*, *C. campinensis* and *P. acnes*) identified in diseased prostate lesions, using two distinct species-specific primers (Fig. 4). Moreover, in order to further corroborate the results, we used two sample cohorts – 'Cohort - 1' (Discovery cohort, used in 16S rRNA amplicon based sequencing) and 'Cohort - 2' (Validation cohort) as described in Table S1. As similar to LEfSe analyses between BPH and PCa biopsy samples, qPCR data further confirmed *Kocuria palustris* and *Cellvibrio mixtus* as BPH specific and *Cupriavidus taiwanensis*, and *Methylobacterium organophilum* as PCa specific bacterial species in both sample cohorts (Figs. 4A-B and 4C-D, respectively). In contrast to 16S rRNA sequencing results, qPCR analyses demonstrated *Prevotella copri*, *Cupriavidus campinensis* and

Propionibacterium acnes were significantly enriched in PCa samples as compared to the BPH group in both sample cohorts (Fig. 4E-G).

Functional prediction of altered microbiome associated with the PCa development: In order to visualize the functional effects resulting from the altered microbial community associated with disease progression in prostate, we employed Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) software [35]. PICRUSt analyses can predict the functional Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways related to the composition of a metagenome, and have been demonstrated to provide a good representation of metagenomic prediction [36]. The LEfSe outputs of KEGG pathways among BPH and PCa tissue biopsy samples identified functions related to starch and sucrose metabolism, galactose metabolism, carbohydrate metabolism, primary immunodeficiency, ubiquitin system, Ion channels, proteasome, phenylpropanoid biosynthesis, electron transfer carriers, glycan degradation and N-Glycan biosynthesis were significantly associated with BPH condition, while pathways such as nitrotoluene degradation, steroid hormone biosynthesis, non-homologous end-joining and primary bile acid biosynthesis were directly linked to PCa development (Fig. 5A). PCoA also demonstrated that the predicted functions of bacterial compositions among BPH and PCa were significantly clustered ($p < 0.05$) (Fig. 5B).

Quantitative real-time PCR analyses detected strong association of multiple human tumor viruses with PCa progression: Studies suggest that a number of human oncogenic viruses including human papilloma viruses (HPVs) and Epstein-Barr virus (EBV) are associated with PCa development[18,20,37]. To evaluate the potential involvement of viral etiology in our samples we designed qPCR primers for seven human tumor viruses including EBV, two high risk HPVs – HPV-16 and HPV-18, hepatitis B virus (HBV), human T-cell leukaemia virus type 1 (HTLV-1), hepatitis C virus (HCV), Kaposi's sarcoma associated herpesvirus (KSHV) and Merkel cell polyomavirus (MCPyV) along with two more human polyomaviruses – JC virus (JCV) and BK virus (BKV) (Table S3). qPCR analyses of cohort-1 with primer set-I comprising of EBV encoded EBNA3A (Gene ID: 3783762), HPV-16 encoded E2 (Gene ID: 1489080) and HPV-18 encoded E6 (Gene ID: 1489088), HBV encoded polymerase (Gene ID: 944565), HCV encoded polyprotein (Gene ID: 951475), KSHV encoded ORF27 (Gene ID: 4961487), HTLV-1 encoded envelope (Gene ID: 1491939), MCPyV encoded VP1 (Gene ID: 10987416), BKV encoded VP1 (Gene ID: 1489515) and JCV encoded Jvgp4 (Gene ID: 1489518) genes demonstrated that only EBV, HPV-16, HPV-18 and HBV are significantly associated with PCa samples as compared to BPH samples (Figs. 6A-D and SA-F). The housekeeping gene human GAPDH gene was utilized as control assuming the genomic segment bearing GAPDH gene remained unaffected in both BPH and PCa samples. A higher negative -ΔCt (average GAPDH Ct value – average target primer Ct value) indicated elevated presence of the virus in the sample as detected by specific primer set targeting specific viral gene. We further validated the association of these four selected tumor viruses in cohort-2 using second set of primers (Fig. 6A-D and Table S3). The results clearly demonstrated that all four tumor viruses were significantly associated with the PCa lesions in comparison to BPH samples (Fig. 6A-D).

Co-occurrence of tumor viruses with microbiome signature linked to BPH and PCa lesions: In order to further corroborate the connection of these tumor viruses with the identified microbial signature associated with BPH and PCa lesions, the co-occurrence and co-exclusion patterns of EBV, HPV-16, HPV-18 and HBV with the most abundant bacterial species identified in LEfSe and qPCR analyses in each category were further investigated (Fig. 7). As depicted in both LEfSe and qPCR analyses, BPH specific bacteria *K. palustris* and *C. mixtus* and PCa specific bacteria *C. taiwanensis* and *M. organophylum* were moderately correlated with each other (r = 0.362 and 0.394, respectively), indicating that they fell into two distinct groups (Fig. 7). Among BPH specific bacteria, interestingly only *K. palustris* was positively correlated with EBV (r = 0.446) and as expected negatively correlated with HPV-16 (r = -0.461) and to lesser extent with HBV (r = -0.263) (Fig. 7). In contrast, BPH specific bacterium *C. mixtus* was only found to be moderately negatively correlated with HPV-16 (r = -0.263) (Fig. 7). Among PCa specific bacteria, while *C. taiwanensis* was positively correlated with HPV-16 (r = 0.411) and to lesser extents with HPV-18 (r = 0.328) and HBV (r = 0.246), *M. organophylum* were positively correlated with EBV (r = 0.456) and to lesser extents with HPV-18 (r = 0.325) and HPV-16 (r = 0.261) (Fig. 7). Among the most abundant bacteria in both BPH and PCa tissue samples, all there species *P. copri*, *C. campinensis* and *P. acnes* were in general found to somewhat positively correlated with the PCa specific bacteria but not with BPH specific bacteria (Fig. 7). Among these three species *P. copri* and *P. acnes* were found to be particularly strongly correlated (r = 0.543) (Fig. 7). In addition, while *P. copri* was positively correlated with both *C. taiwanensis* (r = 0.369) and to a lesser extent *M. organophylum* (r = 0.246), *C. campinensis* and *P. acnes* were correlated with only *M. organophylum* (r = 0.318, and 0.282, respectively) (Fig. 7). While *P. copri* and *P. acnes* were robustly correlated with two tumor viruses – EBV (r = 0.551 and 0.599, respectively) and HPV-18 (r = 0.509 and 0.444, respectively), *C. campinenesis* was weakly correlated with only HPV-18 (r = 0.279) (Fig. 7). Among the 4 tumor viruses, EBV and HPV-18 were positively correlated with each other (r = 0.551), whereas HPV-16 and HBV were grouped together (r = 0.370) (Fig. 7).

Discussion

A growing body of evidence indicated microbial infection as one of the predominant risk factors for PCa development[10,12,20,38]. In addition, several studies derived from 16S rRNA amplicon based sequencing, whole genome shotgun sequencing as well as hybridization based microarray techniques have evidently documented microbiome dysbiosis associated with disease progression[9,19,23,24]. Despite the increasing incidence of PCa [26], to date, there are no reports describing the microbial dysbiosis associated with PCa development among Indian patients. Herein, we identified specific microbial signature including both alteration of tissue specific commensal bacteria along with infection status of several human tumor viruses linked with the pathologic prostate tissue specimens collected from patients in the Eastern region of India.

Multiple studies suggested that an inflammatory microenvironment is involved in the development of PCa precursor lesions that promote tumor initiation[3,39,40]. Although microbial infections are considered as mainstay of chronic inflammation, to date, no single microbe was identified as a specific inducer of

prostatic inflammation or as a direct contributor to PCa development. However, several bacteria are known to induce prostatic inflammation. For example, several species commonly implicated in bacterial prostatitis such as *E. coli* and *Enterococcus faecalis* as well as species responsible for sexually transmitted diseases including *Chlamydia trachomatis*, *Neisseria gonorrhoeae* and *Trichomonas vaginalis* have been shown to induce prostatic inflammation[3,5,11,12]. In addition, *Propionibacterium acnes*, an anaerobic Gram-positive pro-inflammatory bacterium ubiquitously found in sebaceous follicles of the human skin, has been detected with elevated levels in prostate tissue specimens from patients with prostatitis and PCa and can promote prostatic inflammation in both human prostate epithelial cells as well as in animal models[13,14,38,41,42]. Moreover, *P. acnes* infection in prostate epithelial cells resulted in increased cell proliferation and anchorage-independent cell growth[42], indicating the possibility of *P. acnes* as one of the major contributing inflammatory factors for the onset and subsequent development of PCa. In addition, transcriptomic analyses using RNA-Seq data of human prostate samples collected from both Caucasian and Chinese patients demonstrated significant expression of *P. acnes* genes in PCa samples as compared to the matched controls[25]. In agreement to this, our 16S rRNA amplicon sequencing results also demonstrated that *P. acnes* represented as one of the most abundant bacterial species in diseased prostate specimens including both BPH and PCa. Moreover, our qPCR analyses using two different primer sets of both discovery and validation sample cohorts further demonstrated that *P. acnes* was likely to be more associated with PCa samples in comparison to BPH samples. In addition, due to anatomical proximity urinary microbiota has also been shown to influence prostate pathophysiology[43,44]. Various studies profiled urinary microbiota of adult men, which include genera *Corynebacterium, Streptococcus, Veillonella, Prevotella, Anaerococcus, Propionibacterium, Finegoldia, Staphylococcus*, and *Lactobacillus* [45–47]. Another study suggested prevalence of pro-inflammatory bacteria and uropathogens in the urinary tract of PCa patients [43]. In agreement to these previous reports, our results also demonstrated *Prevotella*, *Propionibacterium*, and *Corynebacterium* among the top five genera in diseased prostate lesions including BPH and PCa tissue samples.

Increasing evidence suggested that the gastrointestinal (GI) microbiome controls the efficacy of various cancer treatments, including both chemotherapy and immunotherapy[48–51]. To date, these studies have largely been conducted in animal models and have evidently demonstrated that the compositions of gut microbiome are vital for appropriate functioning of immunotherapeutic agents and modulation of gut microbiome thus could enhance treatment responses. It has been suggested that GI microbiome induced immune modulation is a result of disruption of the intestinal mucosal barrier, which enables interaction between luminal microbiota and mucosa associated lymphoid cells, and thereby a specific composition of the GI microbiome influences this interaction[3]. Although, there are no published data describing the interplay between the GI microbiome and immunotherapy response, similar mechanisms could certainly influence immunotherapy response in case of PCa patients. Thus, in future it would be interesting to investigate to assess the prognostic value of GI microbiome in relation to PCa susceptibility using only diagnostically-confirmed cases of PCa as compared to normal healthy individuals. Of note, recently, Sfanos et al. demonstrated difference in GI microbiome compositions in men undergoing treatment with

androgen deprivation therapies (ADT) commonly used to treat PCa, which may further influence treatment response to oral ADT or to subsequent treatments such as immunotherapy[52]. In functional analyses, an enriched representation of bacterial gene pathways involved in steroid biosynthesis and steroid hormone biosynthesis was observed in the GI microbiome of patients undergoing ADT therapy[52,53]. In agreement to this, our study also demonstrated significant enrichment of steroid hormone biosynthesis pathway in PCa as compared to BPH samples. *Prevotella copri* is a frequent resident of the GI microbiome, and its higher prevalence has been consistently reported in non-Westernized populations[54]. Moreover, *P. copri* was found to be strongly associated with rheumatoid arthritis, a widespread systemic autoimmune disease[55]. Colonization of mice with *P. copri* demonstrated increased susceptibility to chemically induced colitis, signifying a pro-inflammatory function of this organism[55]. Interestingly, our 16S rRNA sequencing results also demonstrated that *P. copri* was the most abundant species in both BPH and PCa lesions. In addition, qPCR analyses showed *P. copri* was significantly associated with PCa samples, indicating potential involvement of GI microbiota with PCa development possibly through modulating inflammatory response.

Intriguing evidence is emerging that indicates a potential association of several human oncogenic viruses with PCa[18,19,21,22,25,56,57]. For example, a number of HPV subtypes, several polyomaviruses such as JCV, BKV and SV-40 as well as herpesviruses including HCMV and EBV were shown to be associated with PCa samples across the globe[18, 20–22,56,58,59]. Moreover, the frequency of co-infection of EBV and high-risk HPV-18 subtype has been shown to be significantly higher in PCa as compared to benign and normal prostate samples[20]. In agreement to this, our results also demonstrated that EBV was strongly correlated with HPV-18 but not with HPV-16 in PCa samples. A recent report by Ishiguro et al. using 30 non-cancer and 182 PCa tissue specimens suggested that HBV and HCV infection are not linked with PCa development in Japanese patients[60]. In contrast, using two different qPCR primer sets our study for the first time demonstrated that HBV infection was significantly associated with PCa development. Importantly, a number of studies suggested that both HBV and HCV infection can promote androgen receptor (AR) signaling in hepatocellular carcinoma (HCC). Intriguingly, the gender preference of HCC differs between HBV and HCV related cases. The male predominance in HBV-linked HCC is significantly higher than that of HCV-linked HCC[61], indicating potential link of AR-signaling and HBV infection. It has been demonstrated that HBV encoded oncoprotein HBx induces AR-responsive gene expression in an androgen concentration dependent manner[62–64]. Targeting the AR signaling axis has been, over decades, the mainstay of PCa therapy[53,65]. Although ADT therapy using several specific AR inhibitors blocks further tumor growth for some patients, most patients develop resistance to the treatment and subsequently develop to castration-resistant PCa (CRPC) associated with poor prognosis[65,66]. Given that AR overexpression enables PCa to progress to castration levels of androgen[53,66]; it would be fascinating to investigate the role of HBV infection emphasizing its oncoprotein HBx in PCa progression. Altogether, it is still unclear whether the co-existence of these tumor viruses along with dysbiosis of tissue specific microbiome acts as a promoter or bystander in PCa development. Nevertheless, in order to improve the treatment strategy as well as development of diagnosis and prognosis markers, it is important, from the

perspective of genetic etiology, to clarify the connection. Our study clearly revealed that *C. taiwanensis*, *M. organophylum* followed by *P. acnes* and *P. copri* were the most correlated bacterial species with the infection status of EBV and HPV-18 and to lesser extents HPV-16 and HBV in PCa tissue specimens.

In sum, our study demonstrated that dysbiosis of tissue specific microbiome is directly linked to prostate health and disease. Future in-depth investigations with larger sample cohorts of different socio-economic and ethnic background are required to discern whether the microbiota and/or their metabolites can be considered as novel biomarkers and therapeutic targets for PCa in Indian scenario. The mechanisms by which the dysbiosis of commensal microbiota as well as infection of oncogenic viruses facilitate tumorigenesis through influencing systemic inflammatory state, affecting hormone levels, metabolism as well as genotoxicity, can provide unique opportunities to explore the microbiome signature for diagnostic, preventive, as well as expansion of current therapeutic strategies against PCa onset and progression.

Materials And Methods

Ethics Statement: The study was approved by the Institutional Review Board of Institute of Post Graduate Medical Education & Research (IPGME&R), Kolkata, India. Written informed consent was obtained from all participants and all methods in this study were performed in accordance with the ethical principles founded in the Declaration of Helsinki.

Subject recruitment and specimen collection: Affected individuals were recruited at Dept. of Urology, IPGME&R, Kolkata. Affected individuals with prostatic lesions suspected to have either BPH or PCa accessible by image-guided biopsy were eligible for inclusion. A standard TRUS-guided 18-core prostate biopsy with subsequent histopathological analysis was carried out in those with abnormal serum PSA level (>4.0 ng/ml). All cases included were newly diagnosed patients without any prior treatment before surgery and diagnosis was based on the histopathological analysis. Individuals with clinical evidence of prostatic lesions receiving radiotherapy and/or undergoing ADT were excluded. In addition, none of the patients reported prior known sexually transmitted infection or a recent history of urinary tract infections. Samples categorized as BPH were being evaluated for suspicion of PCa, but subsequently had a negative biopsy according to histopathological analysis. Digital images of biopsy slides were reviewed by genitourinary oncology pathologist. Samples were collected into two cohorts. Cohort 1 (Discovery Cohort) containing 13 BPH and 33 PCa tissue biopsy samples were subjected to 16S rRNA amplicon sequencing. Cohort 2 (Validation Cohort) containing 15 BPH and 16 PCa tissue biopsy samples together with Cohort 1 were used for validation using real-time qPCR analyses. The median age of the controls was 68 years (range 42–78), whereas the PCa cases had a mean age of 70 years (range 55–79). Clinical characteristics and other information related to both sample cohorts are described in Table S1.

Nucleic acid extraction: A total of 77 formalin-fixed paraffin-embedded (FFPE) tissue samples in both cohorts including 28 BPH and 49 PCa tissue biopsy specimens were received as 10 µm sections on non-charged glass slides. Genomic DNA was isolated according to the standard procedure as previously described[27]. Briefly, both tissue and fluid samples were homogenized in TNE buffer (50 mM Tris-HCl pH

7.4, 100 mM NaCl, 0.1 mM EDTA), mixed with 10% SDS and followed by incubation at 65^0C for 10 min. Next, 1 mg/ml Proteinase K (Thermo Fisher Scientific Inc., Waltham, MA, USA) was added into the solution and incubated at 37^0C for overnight. DNA was purified from the solution using 25:24:1 phenol: Chloroform: isoamyl alcohol (Sigma-Aldrich Corp. St. Louis, MO, USA). Following centrifugation, the aqueous phase containing the purified DNA was subjected to precipitation using cold Ethanol in the presence of sodium acetate at -20^0C for overnight. Subsequently, the precipitated DNA was washed in 70% ethanol and dissolved in 1xTE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA was then stored in -20^0C for future use. The quality and quantity of extracted DNA was determined by agarose gel electrophoresis and the A260/280 ratio using Synergy H1 Multimode Microplate Reader (BioTek Instruments, Inc., VT, USA). Approximately 50 ng and 1 ng of DNA from each sample were used for 16S rRNA amplicon sequencing and real-time qPCR analyses, respectively, as described below.

16S rRNA amplicon sequencing and OTU assignments: For characterization of bacterial populations, taxonomical analysis, and species identification, different hypervariable regions of 16S rRNA gene were amplified and sequenced on an Ion GeneStudio S5 System (Thermo Fisher Scientific Inc., Waltham, MA, USA) using Ion 550 Chip after generation of library by two distinct primer sets as per manufacturer's instructions. While Set I primer was used for amplification of hyper variable regions V2, V4 and V8, set II primer was used for hyper variable regions V3, V6-7 and V9. Ion AmpliSeq Libraries were prepared using an automated Ion Chef system (Thermo Fisher Scientific Inc., Waltham, MA, USA) and subsequently placed in the Ion Chef System for emulsion PCR. The concentration of the amplicons was measured using Qubit fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and the quality was checked using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

To assign taxonomy, each unique sequence was mapped using two comprehensive 16S rRNA reference databases – Greengenes v13.5[28] and the Thermo Fisher Scientific in-house MicroSeq 500 v2013.1.[29] Reads were aligned against the databases using MegaBLAST [30]. The expectation value (E-value) for the searches was set to 0.01, and the max target hits value was set to 100[30]. To assign taxonomy, the minimum alignment percentage of a read to a subject sequence was set to a threshold of 90. A read was assigned to a genus only when the identity score of the sequence alignment was at 97% or higher. For species assignment, the minimum percentage identity of the alignment was set to 99%. The taxonomy distribution counts or abundance derived from the clustered reads was subsequently transformed into the relative abundance of the individual species.

Diversity and bacterial enrichment analyses: For statistical analysis MicrobiomeAnalyst[31], a web-based tool for comprehensive analysis of microbiome data, was utilized. 'Quantitative Insights Into Microbial Ecology' (QIIME v1.9.0)[32] was used to evaluate alpha diversity including Observed Species, Chao1, Shannon indexes. While Observed Species and Chao1 indexes are the indicators of species richness, Shannon index evaluates species diversity. The difference of alpha diversity between groups was evaluated by Wilcoxon Rank-Sum Test (group number = 2) and Kruskal–Wallis test ($n > 2$) using SPSS (version 22).

To compare microbial compositions between different groups, beta diversity was evaluated by calculating weighted UniFrac distances[33] using Bray-Curtis method from the OTU abundance and utilized in Principal Component Analysis (PCoA). PERMANOVA algorithm [34] on weighted UniFrac distance matrices for statistical significance between groups using 999 permutations in QIIME was applied to generate PCoA plots. The Differential abundance analysis between groups was performed using Metastats and *P*-values were adjusted for multiple hypotheses testing using the False Discovery Rate (FDR) based on the Benjamini-Hochberg. The differential abundances of OTUs and specific OTU enrichment between different groups were determined using LEfSe based on Kruskal–Wallis H test. P-value and FDR were adjusted to 0.05. The unique bacterial composition among sample groups was identified using Random Forest classification algorithm [35] within MicrobiomeAnalyst. Galaxy[36], an online tool for metagenomics analysis, was used to build the Cladogram.

Functional pathways prediction: Functional compositions of the bacterial communities among different groups were predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt)[37] according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[38] using STAMP v2.1.3[39]. P value and FDR cut off was adjusted to 0.05 level of significance.

Real-time qPCR analyses: Gene-specific primers for human Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), bacteria including *Kocuria palustris*, *Cellvibrio mixtus, Cupriavidus taiwanensis*, *Methylobacterium organophilum*, *Prevotella copri, Cupriavidus campinensis*, *Propionibacterium acnes*, and oncogenic viruses including EBV, HPV-16, HPV-18, HBV, HCV, HTLV-1, KSHV, MCPyV, BKV, JCV were designed using Primer-BLAST tool in National Center for Biotechnology Information (NCBI) database for real-time qPCR analyses and are listed in Table S3. qPCR primers were obtained from Integrated DNA Technologies, Inc. (Coralville, IA, USA). The optimum primer melting temperature (Tm) was set at 60°C and the maximum GC content was kept at 55%. qPCR analysis was performed using iTaq Universal SYBR Green Supermix (BIO-RAD, CA, USA) in CFX Connect Real-Time PCR detection System (BIO-RAD, CA, USA) with the following thermal profile – one cycle: 95°C for 10 min; 40 cycles: 95°C for 10 s followed by 60°C for 10 s; and finally the dissociation curve at – 95°C for 1 min, 55°C 10 s, and 95°C for 10 s. Unless and otherwise stated, each sample was performed in duplicate and calculation was made using a -ΔCT method to quantify relative abundance compared with human genomic GAPDH control. The -ΔCt values of each sample was plotted using GraphPad Prism 8.0.1 for data output.

Correlation analysis between microbes: Correlation among bacteria and oncogenic viruses were analysed using R software. Spearman's rank test was performed using the ΔCt value of both bacteria and viruses. P value cutoff was adjusted to 0.05 significance level.

Declarations

Acknowledgements

The authors would like to thank all patients for making this study possible. The authors thank all members of AS laboratory for valuable suggestions and discussions of the experiment. A.S. is Wellcome Trust/DBT India Alliance Intermediate Fellow. P.S. and S.M. are the recipients of CSIR-NET and UGC-NET Junior Research Fellowship, respectively.

Funding

This study was supported by grants from Wellcome Trust/DBT India Alliance Intermediate Fellowship research grant [IA/I/14/2/501537], Department of Science and Technology (DST), Govt. of India [CRG/2018/001044] and Science & Technology and Biotechnology, Govt. of West Bengal [1798 (Sanc.)/ST/P/S&T/9G-5/2019] to A.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

P.S. and A.S. wrote the main manuscript text. P.S. and S.M. performed bioinformatic analysis. P.S, A.B. and S.M. performed the experiments. D.K.P. collected the samples. C.D. performed histopathological analysis. A.B. and A.G. performed sampling. A.S. conceived, designed and successfully sought funding for the study. All authors have read and approved the content of this manuscript.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability

The 16S rRNA amplicon sequencing data from this study have been deposited in the NCBI BioProject under accession number PRJNA718483.

References

1. Katongole, P., Sande, O. J., Joloba, M., Reynolds, S. J. & Niyonzima, N. The human microbiome and its link in prostate cancer risk and pathogenesis. *Infect Agent Cancer* **15**, 53, doi:10.1186/s13027-020-00319-2319 [pii] (2020).
2. Hariharan, K. & Padmanabha, V. Demography and disease characteristics of prostate cancer in India. *Indian J Urol* **32**, 103-108, doi:10.4103/0970-1591.174774IJU-32-103 [pii] (2016).
3. Sfanos, K. S., Yegnasubramanian, S., Nelson, W. G. & De Marzo, A. M. The inflammatory microenvironment and microbiome in prostate cancer development. *Nat Rev Urol* **15**, 11-24, doi:nrurol.2017.167 [pii]10.1038/nrurol.2017.167 (2017).
4. Porter, C. M., Shrestha, E., Peiffer, L. B. & Sfanos, K. S. The microbiome in prostate inflammation and prostate cancer. *Prostate Cancer Prostatic Dis* **21**, 345-354, doi:10.1038/s41391-018-0041-110.1038/s41391-018-0041-1 [pii] (2018).

5. Nickel, J. C. & Xiang, J. Clinical significance of nontraditional bacterial uropathogens in the management of chronic prostatitis. *J Urol* **179**, 1391-1395, doi:S0022-5347(07)03121-7 [pii]10.1016/j.juro.2007.11.081 (2008).
6. Fonseca, R. C., Gomes, C. M., Meireles, E. B., Freire, G. C. & Srougi, M. Prostate specific antigen levels following transurethral resection of the prostate. *Int Braz J Urol* **34**, 41-48, doi:IBJUv34n1a6 [pii]10.1590/s1677-55382008000100007 (2008).
7. Mottet, N. *et al.* EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* **79**, 243-262, doi:S0302-2838(20)30769-7 [pii]10.1016/j.eururo.2020.09.042 (2020).
8. Banerjee, J., Mishra, N. & Dhas, Y. Metagenomics: A new horizon in cancer research. *Meta Gene* **5**, 84-89, doi:10.1016/j.mgene.2015.05.005S2214-5400(15)00028-6 [pii] (2015).
9. Massari, F. *et al.* The Human Microbiota and Prostate Cancer: Friend or Foe? *Cancers (Basel)* **11**, doi:cancers11040459 [pii]10.3390/cancers11040459 (2019).
10. Jain, S. *et al.* Escherichia coli, a common constituent of benign prostate hyperplasia-associated microbiota induces inflammation and DNA damage in prostate epithelial cells. *Prostate*, doi:10.1002/pros.24063 (2020).
11. Sutcliffe, S. *et al.* Sexually transmitted infections and prostatic inflammation/cell damage as measured by serum prostate specific antigen concentration. *J Urol* **175**, 1937-1942, doi:S0022-5347(05)00892-X [pii]10.1016/S0022-5347(05)00892-X (2006).
12. Sutcliffe, S. *et al.* Prostate involvement during sexually transmitted infections as measured by prostate-specific antigen concentration. *Br J Cancer* **105**, 602-605, doi:bjc2011271 [pii]10.1038/bjc.2011.271 (2011).
13. Olsson, J. *et al.* Chronic prostatic infection and inflammation by Propionibacterium acnes in a rat prostate infection model. *PLoS One* **7**, e51434, doi:10.1371/journal.pone.0051434PONE-D-12-24563 [pii] (2012).
14. Shinohara, D. B. *et al.* A mouse model of chronic prostatic inflammation using a human prostate cancer-derived isolate of Propionibacterium acnes. *Prostate* **73**, 1007-1015, doi:10.1002/pros.22648 (2013).
15. Cohen, R. J., Shannon, B. A., McNeal, J. E., Shannon, T. & Garrett, K. L. Propionibacterium acnes associated with inflammation in radical prostatectomy specimens: a possible link to cancer evolution? *J Urol* **173**, 1969-1974, doi:S0022-5347(05)60190-5 [pii]10.1097/01.ju.0000158161.15277.78 (2005).
16. Mak, T. N., Yu, S. H., De Marzo, A. M., Bruggemann, H. & Sfanos, K. S. Multilocus sequence typing (MLST) analysis of Propionibacterium acnes isolates from radical prostatectomy specimens. *Prostate* **73**, 770-777, doi:10.1002/pros.22621 (2013).
17. Saha, A., Kaul, R., Murakami, M. & Robertson, E. S. Tumor viruses and cancer biology: Modulating signaling pathways for therapeutic intervention. *Cancer Biol Ther* **10**, 961-978, doi:13923 [pii]10.4161/cbt.10.10.13923 (2010).

18. Smelov, V. *et al.* Detection of DNA viruses in prostate cancer. *Sci Rep* **6**, 25235, doi:srep25235 [pii]10.1038/srep25235 (2016).
19. Banerjee, S. *et al.* Microbiome signatures in prostate cancer. *Carcinogenesis* **40**, 749-764, doi:5362026 [pii]10.1093/carcin/bgz008 (2019).
20. Whitaker, N. J. *et al.* Human papillomavirus and Epstein Barr virus in prostate cancer: Koilocytes indicate potential oncogenic influences of human papillomavirus in prostate cancer. *Prostate* **73**, 236-241, doi:10.1002/pros.22562 (2013).
21. Samanta, M., Harkins, L., Klemm, K., Britt, W. J. & Cobbs, C. S. High prevalence of human cytomegalovirus in prostatic intraepithelial neoplasia and prostatic carcinoma. *J Urol* **170**, 998-1002, doi:10.1097/01.ju.0000080263.46164.97S0022-5347(05)63295-8 [pii] (2003).
22. Das, D., Wojno, K. & Imperiale, M. J. BK virus as a cofactor in the etiology of prostate cancer in its early stages. *J Virol* **82**, 2705-2714, doi:JVI.02461-07 [pii]10.1128/JVI.02461-07 (2008).
23. Cavarretta, I. *et al.* The Microbiome of the Prostate Tumor Microenvironment. *Eur Urol* **72**, 625-631, doi:S0302-2838(17)30250-6 [pii]10.1016/j.eururo.2017.03.029 (2017).
24. Yow, M. A. *et al.* Characterisation of microbial communities within aggressive prostate cancer tissues. *Infect Agent Cancer* **12**, 4, doi:10.1186/s13027-016-0112-7112 [pii] (2017).
25. Chen, Y. & Wei, J. Identification of Pathogen Signatures in Prostate Cancer Using RNA-seq. *PLoS One* **10**, e0128955, doi:10.1371/journal.pone.0128955PONE-D-14-42944 [pii] (2015).
26. Jain, S., Saxena, S. & Kumar, A. Epidemiology of prostate cancer in India. *Meta Gene* **2**, 596-605, doi:10.1016/j.mgene.2014.07.007S2214-5400(14)00048-6 [pii] (2014).
27. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069-5072, doi:72/7/5069 [pii]10.1128/AEM.03006-05 (2006).
28. Fontana, C., Favaro, M., Pelliccioni, M., Pistoia, E. S. & Favalli, C. Use of the MicroSeq 500 16S rRNA gene-based sequencing for identification of bacterial isolates that commercial automated systems failed to identify correctly. *J Clin Microbiol* **43**, 615-619, doi:43/2/615 [pii]10.1128/JCM.43.2.615-619.2005 (2005).
29. Dhariwal, A. *et al.* MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* **45**, W180-W188, doi:3760191 [pii]10.1093/nar/gkx295 (2017).
30. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**, 169-172, doi:ismej2010133 [pii]10.1038/ismej.2010.133 (2010).
31. Kelly, B. J. *et al.* Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* **31**, 2461-2468, doi:btv183 [pii]10.1093/bioinformatics/btv183 (2015).
32. Goodrich, J. K. *et al.* Conducting a microbiome study. *Cell* **158**, 250-262, doi:S0092-8674(14)00864-2 [pii]10.1016/j.cell.2014.06.037 (2014).

33. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60, doi:gb-2011-12-6-r60 [pii]10.1186/gb-2011-12-6-r60 (2011).
34. Poretsky, R., Rodriguez, R. L., Luo, C., Tsementzi, D. & Konstantinidis, K. T. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* **9**, e93827, doi:10.1371/journal.pone.0093827PONE-D-14-00837 [pii] (2014).
35. Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**, 814-821, doi:nbt.2676 [pii]10.1038/nbt.2676 (2013).
36. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30, doi:gkd027 [pii]10.1093/nar/28.1.27 (2000).
37. Pascale, M. *et al.* Is human papillomavirus associated with prostate cancer survival? *Dis Markers* **35**, 607-613, doi:10.1155/2013/735843 (2013).
38. Sfanos, K. S. *et al.* A molecular analysis of prokaryotic and viral DNA sequences in prostate tissue from patients with prostate cancer indicates the presence of multiple and diverse microorganisms. *Prostate* **68**, 306-320, doi:10.1002/pros.20680 (2008).
39. Kwon, O. J., Zhang, L., Ittmann, M. M. & Xin, L. Prostatic inflammation enhances basal-to-luminal differentiation and accelerates initiation of prostate cancer with a basal cell origin. *Proc Natl Acad Sci U S A* **111**, E592-600, doi:1318157111 [pii]10.1073/pnas.1318157111 (2014).
40. Palapattu, G. S. *et al.* Prostate carcinogenesis and inflammation: emerging insights. *Carcinogenesis* **26**, 1170-1181, doi:bgh317 [pii]10.1093/carcin/bgh317 (2005).
41. Sfanos, K. S. & Isaacs, W. B. An evaluation of PCR primer sets used for detection of Propionibacterium acnes in prostate tissue samples. *Prostate* **68**, 1492-1495, doi:10.1002/pros.20820 (2008).
42. Fassi Fehri, L. *et al.* Prevalence of Propionibacterium acnes in diseased prostates and its inflammatory and transforming activity on prostate epithelial cells. *Int J Med Microbiol* **301**, 69-78, doi:S1438-4221(10)00096-2 [pii]10.1016/j.ijmm.2010.08.014 (2011).
43. Shrestha, E. *et al.* Profiling the Urinary Microbiome in Men with Positive versus Negative Biopsies for Prostate Cancer. *J Urol* **199**, 161-171, doi:S0022-5347(17)77253-9 [pii]10.1016/j.juro.2017.08.001 (2018).
44. Yu, H. *et al.* Urinary microbiota in patients with prostate cancer and benign prostatic hyperplasia. *Arch Med Sci* **11**, 385-394, doi:10.5114/aoms.2015.5097025020 [pii] (2015).
45. Dong, Q. *et al.* The microbial communities in male first catch urine are highly similar to those in paired urethral swab specimens. *PLoS One* **6**, e19709, doi:10.1371/journal.pone.0019709PONE-D-11-04657 [pii] (2011).
46. Fouts, D. E. *et al.* 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, doi:1479-5876-10-174 [pii]10.1186/1479-5876-10-174 (2012).

47. Nelson, D. E. *et al.* Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PLoS One* **5**, e14116, doi:10.1371/journal.pone.0014116 (2010).
48. Iida, N. *et al.* Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* **342**, 967-970, doi:342/6161/967 [pii]10.1126/science.1240527 (2013).
49. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350**, 1084-1089, doi:science.aac4255 [pii]10.1126/science.aac4255 (2015).
50. Vetizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350**, 1079-1084, doi:science.aad1329 [pii]10.1126/science.aad1329 (2015).
51. ;Viaud, S. *et al.* The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* **342**, 971-976, doi:342/6161/971 [pii]10.1126/science.1240537 (2013).
52. Sfanos, K. S. *et al.* Compositional differences in gastrointestinal microbiota in prostate cancer patients treated with androgen axis-targeted therapies. *Prostate Cancer Prostatic Dis* **21**, 539-548, doi:10.1038/s41391-018-0061-x10.1038/s41391-018-0061-x [pii] (2018).
53. Tiwari, R. *et al.* Androgen deprivation upregulates SPINK1 expression and potentiates cellular plasticity in prostate cancer. *Nat Commun* **11**, 384, doi:10.1038/s41467-019-14184-010.1038/s41467-019-14184-0 [pii] (2020).
54. Tett, A. *et al.* The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host Microbe* **26**, 666-679 e667, doi:S1931-3128(19)30427-5 [pii]10.1016/j.chom.2019.08.018 (2019).
55. Scher, J. U. *et al.* Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *Elife* **2**, e01202, doi:2/0/e01202 [pii]10.7554/eLife.01202 (2013).
56. Lin, Y. *et al.* Human papillomavirus 16 or 18 infection and prostate cancer risk: a meta-analysis. *Ir J Med Sci* **180**, 497-503, doi:10.1007/s11845-011-0692-6 (2011).
57. Schlaberg, R., Choe, D. J., Brown, K. R., Thaker, H. M. & Singh, I. R. XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. *Proc Natl Acad Sci U S A* **106**, 16351-16356, doi:0906922106 [pii]10.1073/pnas.0906922106 (2009).
58. Deeb, K. K. *et al.* Identification of an integrated SV40 T/t-antigen cancer signature in aggressive human breast, prostate, and lung carcinomas with poor prognosis. *Cancer Res* **67**, 8065-8080, doi:67/17/8065 [pii]10.1158/0008-5472.CAN-07-1515 (2007).
59. Anzivino, E. *et al.* High Frequency of JCV DNA Detection in Prostate Cancer Tissues. *Cancer Genomics Proteomics* **12**, 189-200, doi:12/4/189 [pii] (2015).
60. Ishiguro, H. *et al.* Detection of serum hepatitis B Virus antigen and hepatitis C virus antibody from prostate cancer patients in Japan. *Integr Mol Med* **4**, 1-3, doi: 10.15761/IMM.1000271 (2017).
61. El-Serag, H. B. & Rudolph, K. L. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* **132**, 2557-2576, doi:S0016-5085(07)00799-8 [pii]10.1053/j.gastro.2007.04.061 (2007).

62. Chiu, C. M. *et al.* Hepatitis B virus X protein enhances androgen receptor-responsive gene expression depending on androgen level. *Proc Natl Acad Sci U S A* **104**, 2571-2578, doi:0609498104 [pii]10.1073/pnas.0609498104 (2007).
63. Yang, W. J. *et al.* Hepatitis B virus X protein enhances the transcriptional activity of the androgen receptor through c-Src and glycogen synthase kinase-3beta kinase pathways. *Hepatology* **49**, 1515-1524, doi:10.1002/hep.22833 (2009).
64. Zheng, Y., Chen, W. L., Ma, W. L., Chang, C. & Ou, J. H. Enhancement of gene transactivation activity of androgen receptor by hepatitis B virus X protein. *Virology* **363**, 454-461, doi:S0042-6822(07)00085-2 [pii]10.1016/j.virol.2007.01.040 (2007).
65. Karantanos, T., Corn, P. G. & Thompson, T. C. Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene* **32**, 5501-5511, doi:onc2013206 [pii]10.1038/onc.2013.206 (2013).
66. Cai, Z., Chen, W., Zhang, J. & Li, H. Androgen receptor: what we know and what we expect in castration-resistant prostate cancer. *Int Urol Nephrol* **50**, 1753-1764, doi:10.1007/s11255-018-1964-010.1007/s11255-018-1964-0 [pii] (2018).
67. Ma, X. *et al.* The Microbiome of Prostate Fluid Is Associated With Prostate Cancer. *Front Microbiol* **10**, 1664, doi:10.3389/fmicb.2019.01664 (2019).

Figures

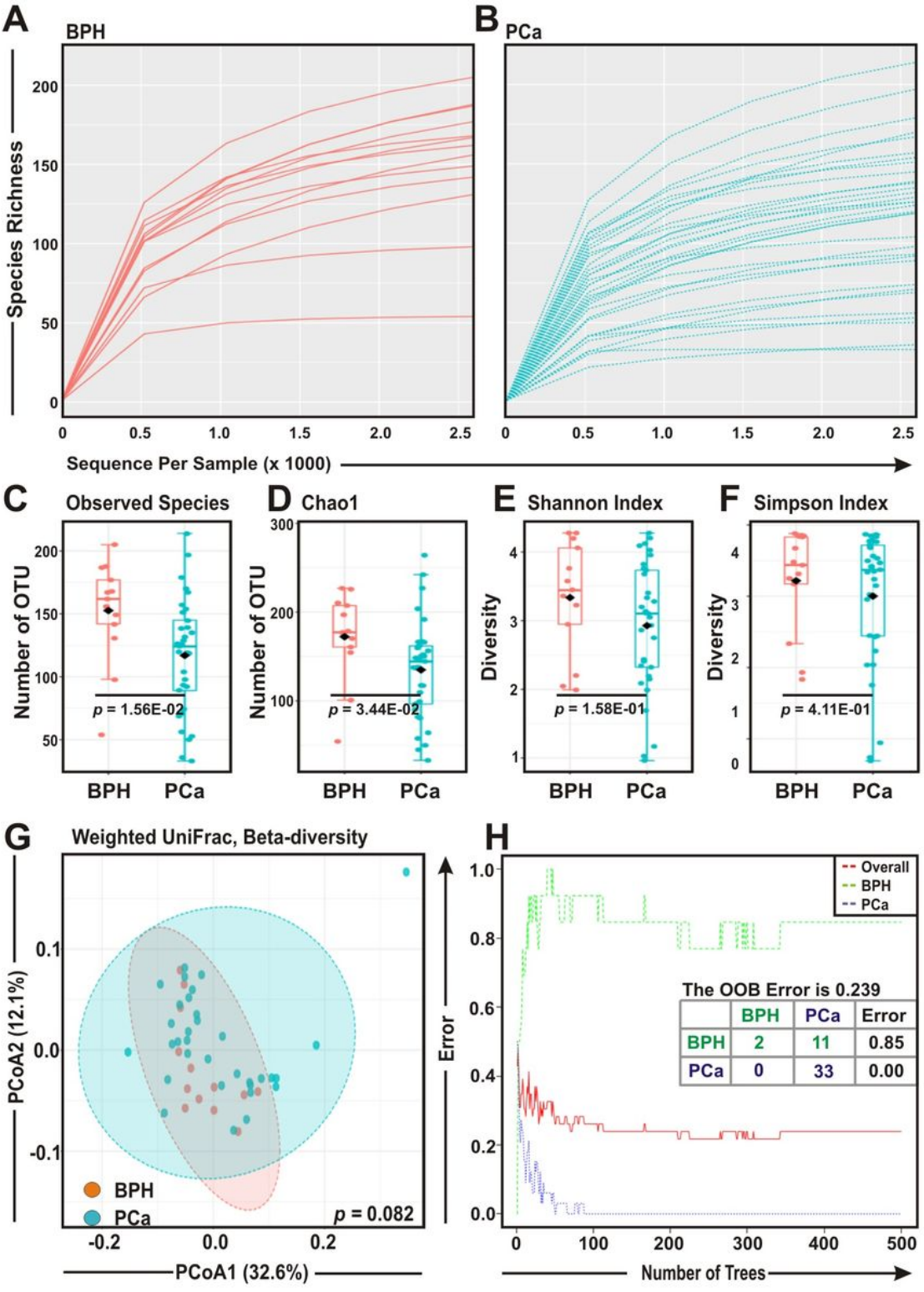

Figure 1

Comparison of prostate microbial ecology among benign prostatic hyperplasia (BPH) and prostate cancer (PCa) specimens. (A-B) Rarefaction analysis of bacterial 16S rRNA amplicon sequences of (A) 13 BPH (red) and (B) 33 PCa (blue) tissue biopsy samples. Each line represents one sample. (C-F) Box-Whisker plots of (C) Observed species, (D) Chao 1, (E) Shannon Index and (F) Simpson Index, respectively among BPH and PCa samples. (G) Principal Component Analysis (PCoA) plot based on weighted UniFrac

distance matrices with respect to the bacterial abundance and composition among BPH and PCa tissue biopsy specimens. Axis 1 (PCoA1): 32.6% of variation explained. Axis 2 (PCoA2): 12.1% of variation explained. (H) The error plots identified from random forest classification analyses of BPH and PCa samples. Red-line indicates the overall species present in all samples including. While green-line indicates the distinct species present in BPH samples, purple-line indicates the specific species present in PCa lesions.

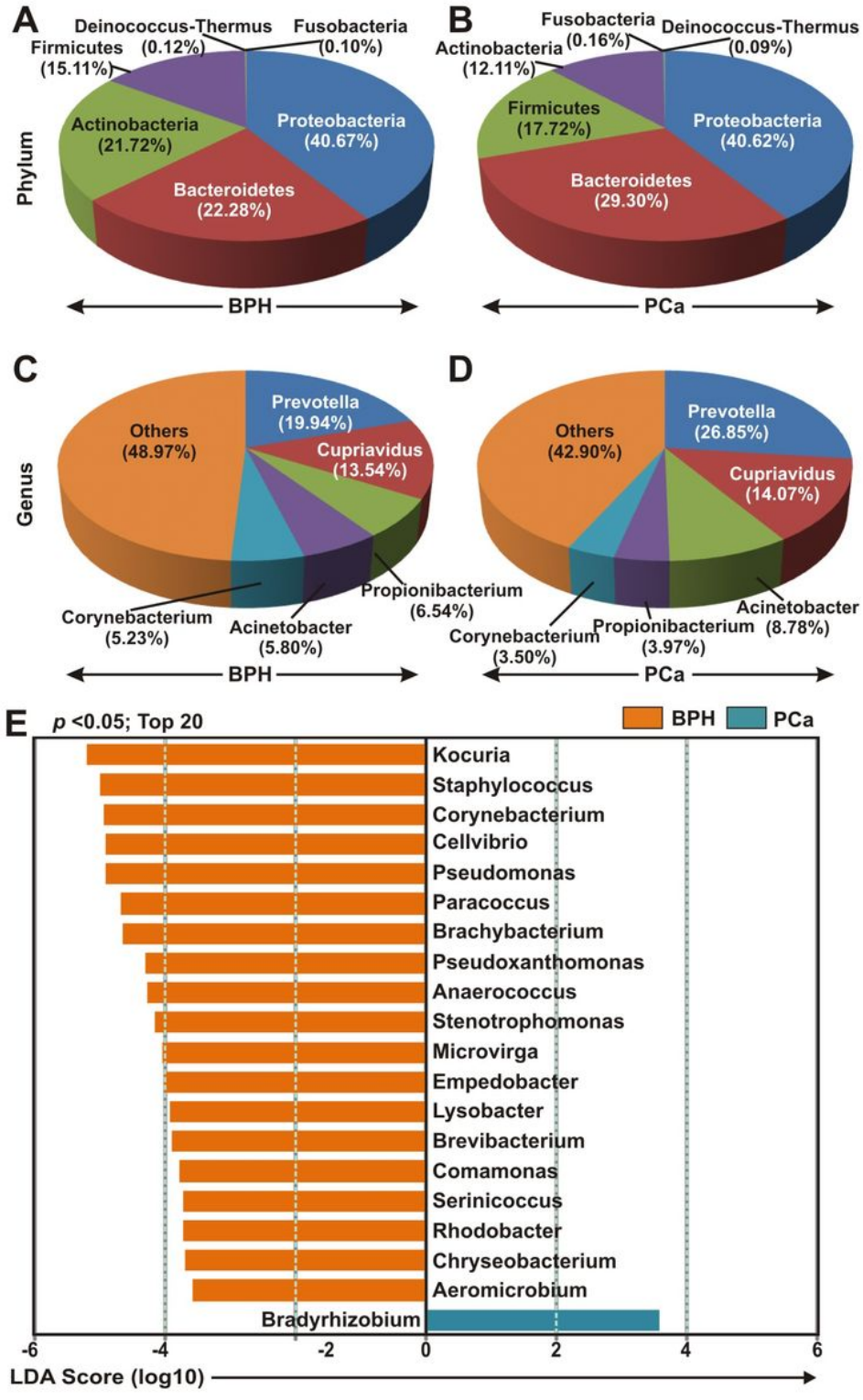

Figure 2

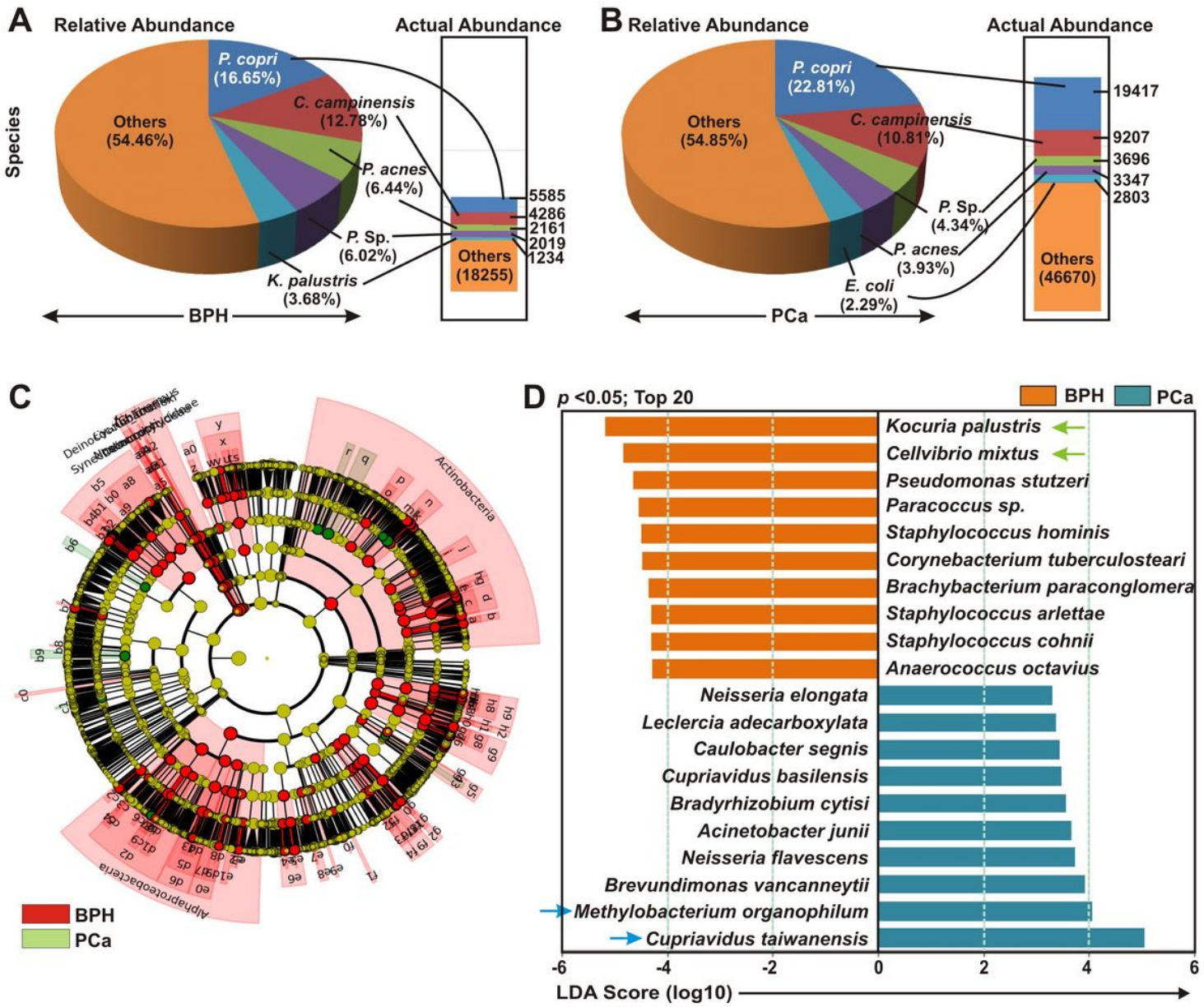

Figure 3

Cladogram and enrichment analysis of specific bacterial species among BPH and PCa lesions. (A-B) Relative and actual abundance of bacterial communities at the species levels in (A) BPH and (B) PCa tissue samples. (C) Cladogram derived from LEfSe analyses showing the taxonomic differences among BPH (red) and PCa (green) tissue biopsy specimens. (D) Association of specific microbiota taxa (top 20

bacterial species) with BPH and PCa samples by LEfSe analyses. Orange indicates taxa enriched in BPH and green indicates taxa enriched in PCa samples.

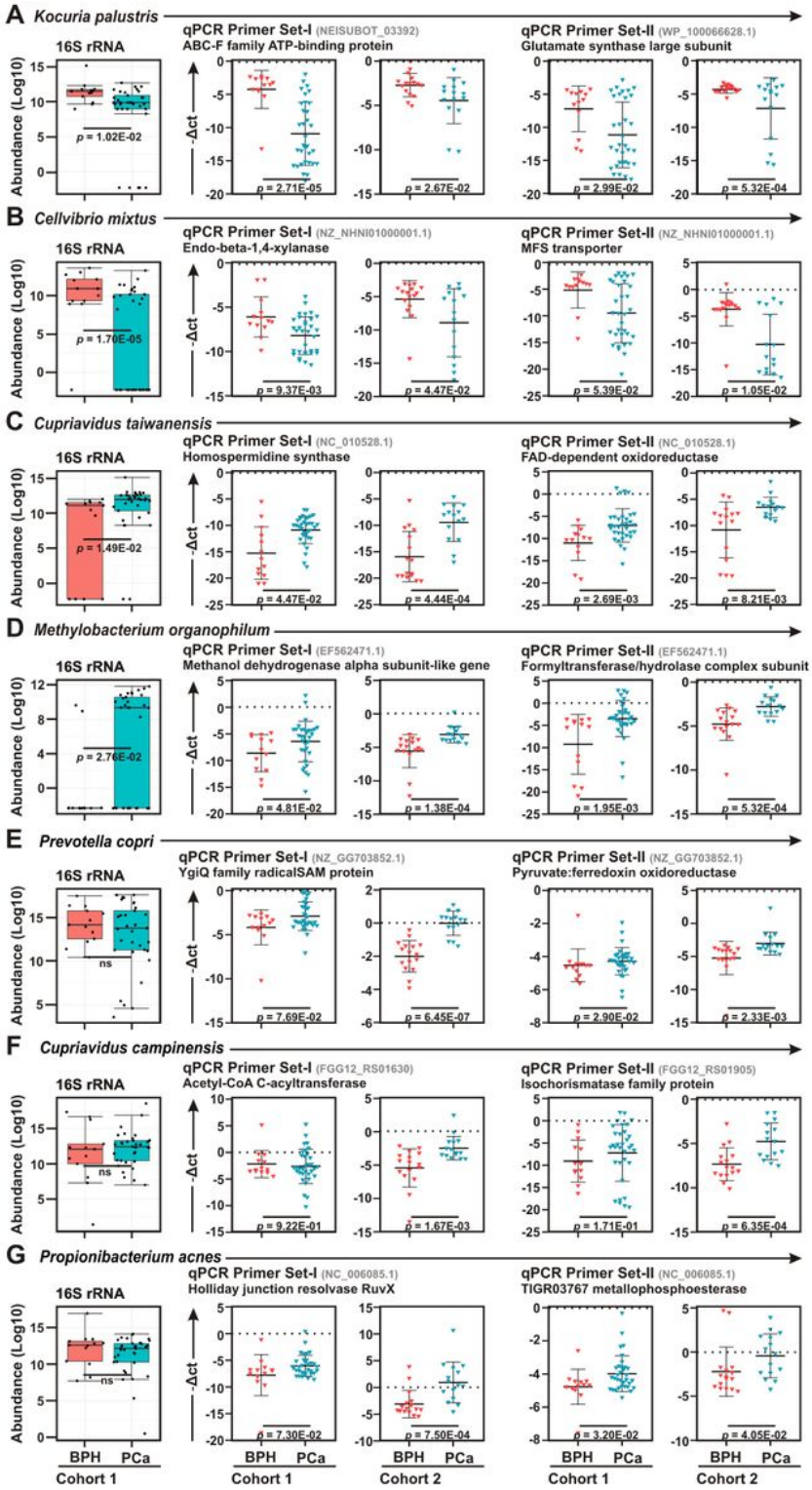

Figure 4

Real-time qPCR confirmation of bacterial species among BPH and PCa lesions. (A-G) Box Whisker plots (left panels) of species enrichment analysis of (A-D) top two bacterial species identified in LEfSe analysis

as described in Fig. 3D and (E-G) top three most abundant bacterial species as described in Fig. 3A-B among BPH (orange) and PCa (green) tissue biopsy lesions of Cohort-1 16S rRNA sequencing results. (A-G) Relative qPCR analyses (right panels) of (A) *Kocuria palustris*, (B) *Cellvibrio mixtus*, (C) *Cupriavidus taiwanensis*, (D) *Methylobacterium organophilum* (E) *Prevotella copri*, (F) *Cupriavidus campinensis* and (G) *Propionibacterium acnes* among BPH (orange) and PCa (green) samples of both cohort-1 and cohort-2 using two sets of primers. Specific gene primer is indicated on the top of each graph. Primers were designed using Primer-BLAST tool of NCBI web-portal. PCR calculation was performed by -ΔCT method to quantify relative abundance of each bacterium using human genomic GAPDH as control. The -ΔCt values of each sample were plotted using GraphPad Prism 8.0.1.

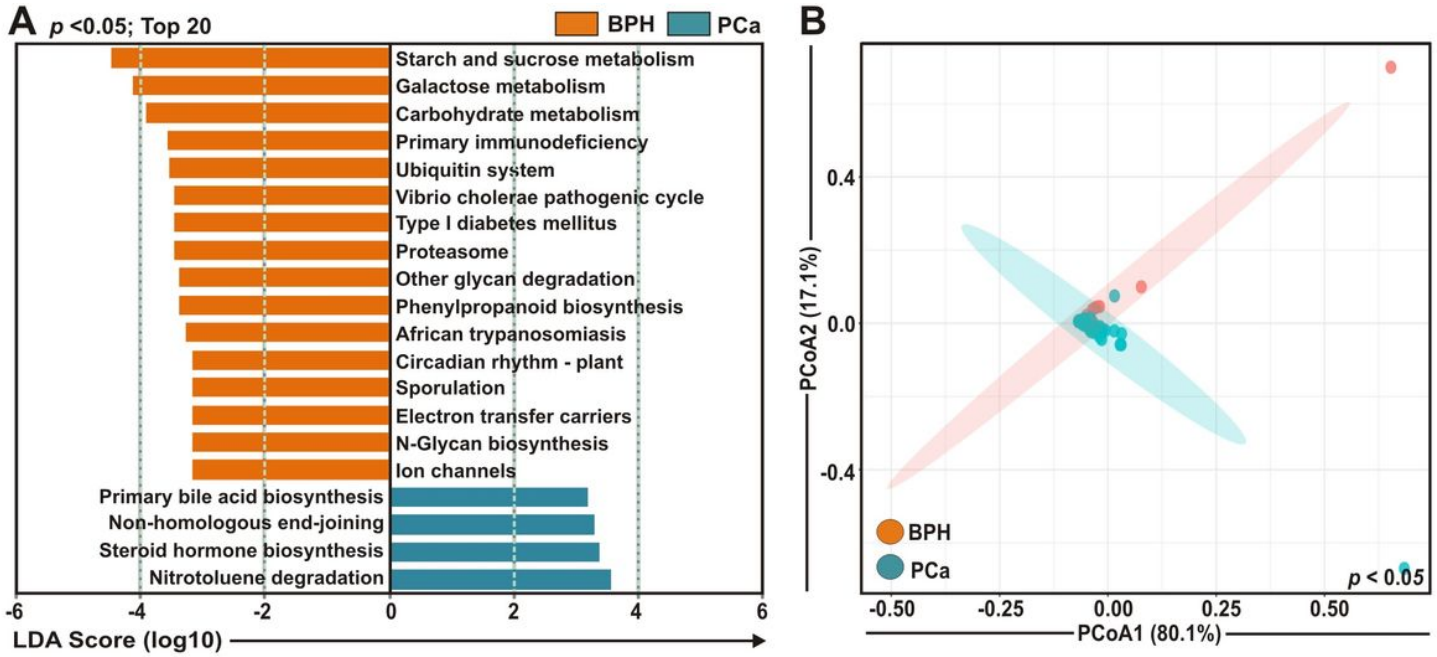

Figure 5

Predicted functional pathways associated with BPH and PCa lesions. (A) LEfSe analyses of potential functional pathways associated with BPH (orange) and PCa (green) tissue biopsy samples. Functional compositions of the bacterial species among different samples groups were predicted using PICRUSt according to the KEGG database. P value and FDR cut off was adjusted to 0.05 level of significance. (B) Principal Component Analysis (PCoA) plot comparing the bacterial functions associated with BPH and PCa tissue biopsy samples. Axis 1 (PCoA1): 80.1% of variation explained. Axis 2 (PCoA2): 17.1% of variation explained.

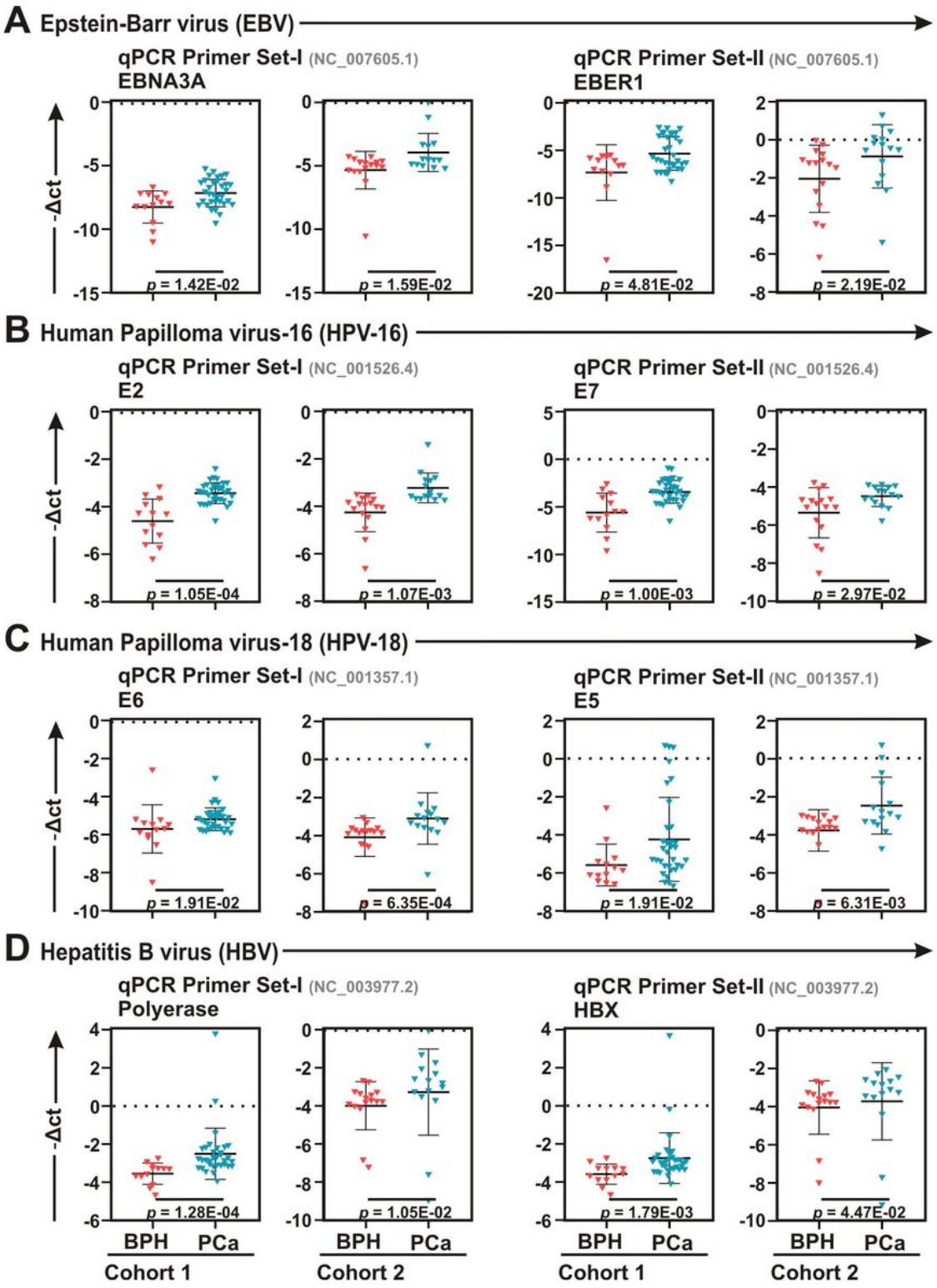

Figure 6

Real-time qPCR analyses of human tumor viruses with two different sets of primers in BPH and PCa lesions. (A-D) Comparative qPCR data of (A) EBV, (B) HPV-16, (C) HPV-18 and (D) HBV among BPH and PCa samples of both cohort-1 and cohort-2 using two sets of primers. Specific gene primer is indicated on the top of each graph. Primers were designed using Primer-BLAST tool in NCBI database. PCR calculation was performed by -ΔCT method to quantify relative abundance of each tumor virus using

human genomic GAPDH as control. The -ΔCt values of each sample were plotted using GraphPad Prism 8.0.1.

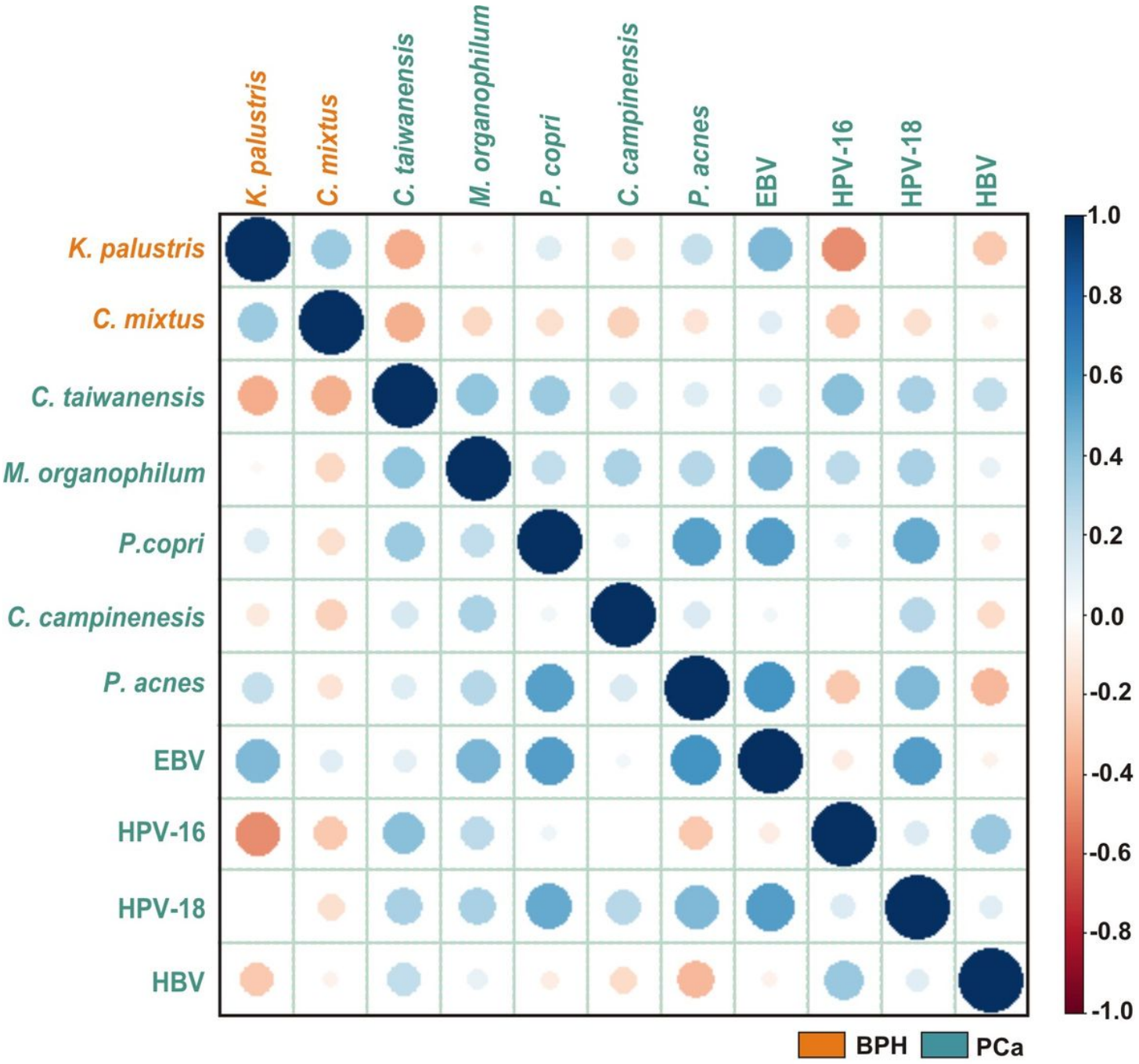

Figure 7

Co-occurrence and co-exclusion analyses between most abundant bacteria and human oncogenic viruses in normal and diseased prostate lesions. Pearson correlations among EBV, HPV-16, HPV-18 and HBV with the top 2 most abundant bacterial species identified in LEfSe and qPCR analyses in each group of normal, BPH, PCa tissue biopsy samples along with normal healthy controls were calculated and analyzed. Correlation values range from -1.0 (red) to +1.0 (blue). Orange: BPH specific; Green: PCa specific.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Fig.S1ScientificReports.tif
- Fig.S2ScientificReports.tif
- TableS1ScientificReports.docx
- TableS2ScientificReports.docx
- TableS3ScientificReports.docx