Carcinogenesis, 2019, Vol. 40, No. 6, 749–764

doi:10.1093/carcin/bgz008
Advance Access Publication February 22, 2019
Original Article

ORIGINAL ARTICLE

Microbiome signatures in prostate cancer

Sagarika Banerjee[1], James C.Alwine[2], Zhi Wei[3], Tian Tian[3], Natalie Shih[4], Colin Sperling[5], Thomas Guzzo[5], Michael D.Feldman[4] and Erle S.Robertson[1*]

[1]Department of Otorhinolaryngology-Head and Neck Surgery and [2]Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [3]Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA and [4]Department of Pathology and Laboratory Medicine and [5]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

*To whom correspondence should be addressed. Tel: +1 215 746 0114; Fax: +1 215 898 9557; Email: erle@pennmedicine.upenn.edu

Abstract

We have established a microbiome signature for prostate cancer using an array-based metagenomic and capture-sequencing approach. A diverse microbiome signature (viral, bacterial, fungal and parasitic) was observed in the prostate cancer samples compared with benign prostate hyperplasia controls. Hierarchical clustering analysis identified three distinct prostate cancer-specific microbiome signatures. The three signatures correlated with different grades, stages and scores of the cancer. Thus, microbiome signature analysis potentially provides clinical diagnosis and outcome predictions. The array data were validated by PCR and targeted next-generation sequencing (NGS). Specific NGS data suggested that certain viral genomic sequences were inserted into the host somatic chromosomes of the prostate cancer samples. A randomly selected group of these was validated by direct PCR and sequencing. In addition, PCR validation of *Helicobacter* showed that *Helicobacter* cagA sequences integrated within specific chromosomes of prostate tumor cells. The viral and *Helicobacter* integrations are predicted to affect the expression of several cellular genes associated with oncogenic processes.

Introduction

Prostate cancer is the most common cancer in males, accounting for 10% of the predicted new US cancer cases in 2019 (1). Apart from other risk factors, chronic inflammation has been correlated with the onset of prostate cancers (2,3). The possible sources of such infection include microorganisms (4–6). For example, gram-positive bacteria such as *Propionibacterium acnes* have been found to be associated with prostate cancer tissues in several studies (4,5). In addition, specific viruses such as human papillomaviruses (HPV), polyomaviruses (BK, JC and SV40) and herpes viruses [human cytomegalovirus (HCMV), Epstein Barr virus (EBV)] are reported to be associated with prostate tumors (7–11). Given the evidence for the presence of associated viral and other microbial agents, we used an array-based metagenomic analysis to define the microbiome of prostate tumors compared with non-cancerous prostate tissue. Using the same metagenomic approach, we have previously reported the microbiome signatures of different breast cancers (12,13), oral cancer and ovarian cancers (14,15).

Previous studies reporting viral and bacterial agents associated with prostate tumors have predominately used PCR-based targeted detection (7,10,11,16) and pyrosequencing (17). One study used RNA-seq to identify pathogens and their integrations within the host genome of prostate cancer (18). This study detected few microorganisms and no viral or bacterial integration in host somatic chromosomes (18). However, conventional RNA-seq methods produced enormous amounts of sequencing data that can be difficult to analyze and even harder to locate microbial sequences within the overwhelming amount of human sequences. Our pan-pathogen array (PathoChip) was designed from NCBI GenBank sequences and presently contains over 6000 accessions and provides the ability to rapidly detect all known viruses, as well as all human pathogenic bacteria, fungi and parasites within the RNA and DNA extracted from any biological material (19).

In the present study, we defined the microbiome (viral, bacterial, fungal and parasitic) signatures associated with

Received: March 20, 2018; **Revised:** November 21, 2018; **Accepted:** February 1, 2019

Abbreviations

BPH	benign prostatic hyperplasia
FFPE	formalin-fixed paraffin embedded
HHV	human herpesvirus
HPV	human papillomaviruses
IPA	Ingenuity Pathway Analysis
KSHV	Kaposi sarcoma associated herpesvirus
MMTV	mouse mammary tumor virus
NCAM1	neural cell adhesion molecule 1
NGS	next-generation sequencing
PPP1R9A	protein phosphatase 1 regulatory subunit 9A
WTA	whole-genome and transcriptome amplification

prostate cancer, which may provide diagnostic and prognostic information that may guide treatment strategies. Three distinct microbiome signatures were defined that show correlations with clinical diagnostic data. We further determined that viral and bacterial sequences were integrated into the tumor cell somatic chromosomes, suggesting increased recombinatorial activity in the tumor cells. Many of these integrations are predicted to affect host genes associated with oncogenic activities in tumor cells.

Materials and methods

All the experiments were performed according to relevant guidelines and regulations as needed and according to all the licensing and approvals by institutional committees at Perelman School of Medicine, University of Pennsylvania.

PathoChip design

The details of the PathoChip array have been described previously (12,19). It comprises 60 000 probes representing sequenced viruses and microorganisms in GenBank. The proprietary arrays were manufactured as SurePrint glass slide microarrays (Agilent Technologies Inc.), containing eight replicate arrays per slide. Each probe is a 60 nucleotide DNA oligomer that targets genomic regions of viruses, prokaryotic and eukaryotic microorganisms. The PathoChip technology, combined with PCR and next-generation sequencing (NGS), is a valuable strategy for detecting and identifying pathogens in human cancers and infection-related pathologies (12,14,15,19). Accession annotations are available in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) (12).

Sample preparation and microarray processing

Fifty formalin-fixed paraffin-embedded (FFPE) prostate adenocarcinoma samples were received as 10 µm sections on non-charged glass slides. All samples were taken from patients who underwent prostatectomy (Supplementary Table S1, available at *Carcinogenesis* Online). As controls, 15 FFPE samples were obtained from patients with benign prostatic hyperplasia (BPH) who underwent transurethral resection of the prostate (Supplementary Table S1, available at *Carcinogenesis* Online). These samples were provided as paraffin rolls. All the samples were obtained from the Abramson Cancer Center's Tumor Tissue and Biosample Core. All FFPE blocks had been stored at room temperature. Nucleic acids retrieved from such blocks, even those aged >40 years, have been shown to be used for molecular analysis (20–22). All samples were de-identified, and thus there was no requirement for informed consent in accordance with Federal and University guidelines. Consequently, we obtained minimal clinical information for these samples, including age of the patients, grade and stage of the tumor (Supplementary Table S1, available at *Carcinogenesis* Online). Both the tumor and control tissues were prepared, examined and verified, by Drs Natalie Shih and Michael Feldman, pathologists at the Hospital of the University of Pennsylvania. Utmost care was taken during the procurement and handling of the samples, and during the process of PathoChip screening to minimize the possibility of contamination.

As described previously, our screening utilized both DNA and RNA extracted from the FFPE samples (12,14,15,19). DNA and RNA were extracted in parallel from rolls or mounted sections of each FFPE sample. The quality of extracted nucleic acids was determined by agarose gel electrophoresis and the A260/280 ratio. The extracted RNA and DNA samples were subjected to whole-genome and transcriptome amplification (referred here as WTA) using the TransPlex Complete Whole Transcriptome Amplification Kit (Sigma-Aldrich, St. Louis, MO) using 50 ng each of RNA and DNA as input. A total of 50 arrays were used to screen the 50 prostate cancer samples, and the control samples were pooled in groups of five samples for screening each array. The WTA products were analyzed by agarose gel electrophoresis and showed an amplicon size range of 200–400 bp, with no contamination in the non-template control used during WTA. Human reference RNA and DNA were also extracted from the human B cell line, BJAB (obtained from ATCC and cultured in the laboratory for <6 months); 15 ng of each was used for WTA. The cellular DNA/RNA provides a reference to compensate for dye bias. The WTA products were purified (PCR purification kit; Qiagen, Germantown, MD), and 1 µg of the amplified products from the cancer and control tissues was labeled with Cy3 and that from the human reference was labeled with Cy5 (SureTag labeling kit, Agilent Technologies, Santa Clara, CA). The labeled cDNA/DNAs were purified, and the efficiencies of labeling were determined by measuring absorbance at 550 nm (for Cy3) and 650 nm (for Cy5). The labeled samples (Cy3 plus Cy5) were hybridized to the PathoChip as described previously (12,14,15,19). The hybridization cocktail [Comparative Genomic Hybridization (CGH) blocking agent and hybridization buffer] was added to each of the labeled test samples (Cy3) mixed with reference (Cy5), denatured and hybridized to the arrays in eight-chamber gasket slides. The slides were incubated at 65°C with rotation, washed and then scanned for visualization using an Agilent SureScan G4900DA array scanner (19).

Microarray data extraction and statistical analysis

The microarray data extraction and analyses have been described previously (12,14,15,19). The raw data from the microarray images were extracted using Agilent Feature Extraction software. We used the R program for normalization and data analyses (23). The microarray screen data are available in Gene Expression Omnibus (Accession No. GSE111648). We calculated scale factor using the signals of green and red channels for human probes. Scale factors are the sum of green and sum of red signal ratios $[\Sigma(g)/\Sigma(r)]$ of human probes. Then, we used scale factors to obtain normalized signals for all other probes. For all probes except human probes, normalized signal is log2 transformed of green signals/scale factors modified red signals (log2 g – scale factor * log2 r). On the normalized signals, t-test is applied to select probes significantly present in cancer samples by comparing cancer samples versus controls and to select probes significantly present in the prostate cancer samples versus the controls. The cutoff for significant detections in cancers versus the controls was log2 fold change > 1 and adjusted P value (with multiple testing corrections) < 0.05. We assumed a two-sample one-sided t-test and set true difference to be 1.27 and SD to be 0.63, based on our PathoChip array data. Then, under the nominal significant level of 0.03 (corresponding to adjusted $P < 0.05$ for the prostate cancer data set), we calculated the power to be 0.92. Prevalence was calculated by counting the number of cancer cases with hybridization signal greater than the average signal of dark corner or negative control probes, and represented as a percentage.

Analyses at the individual probe level (for both specific and conserved probes) and at the family (for viruses) or genera (for bacteria, fungi and parasite) level, taking into account all the probes per family or genera, were performed. We ranked the microbial detections based on their total hybridization signal (sum of significant hybridization signal per family or genera) and prevalence.

The cancer samples were also subjected to unsupervised hierarchical clustering, based on the detection of microbial signatures in the samples (average hybridization signal per viral family or microbial genus), using the R program (Euclidean distance, complete linkage, non-adjusted values) (23,24), and the clusters were validated by Calinski and Harabasz index, which is implemented in R package as NbClust (25). Calinski and

Harabasz index is a cluster index that maximizes intercluster distances and minimizes intracluster distances. We calculated the possible cluster solution that would maximize the index values to achieve the best clustering of the data. The significant differences between the clusters observed by these methods were determined using two-sided *t*-test. The ANOVA test was carried out to find the common signatures significantly present in all the clusters.

Using the limited clinical data provided under HIPAA regulations for de-identified patient samples, we determined the trends in different grades or stages of the cancer that might correlate with specific prostate cancer microbiome signatures. The Gleason system is used to grade prostate cancers based on the number of cells in the cancer tissue that resemble normal prostate tissue under the microscope; if the cancerous tissue looks much like normal prostate tissue, a grade of 1 is assigned; if the cancer cells and their growth patterns look very abnormal, a grade of 5 is assigned; Grades 2 through 4 have features in between these extremes. Because prostate cancers often have areas with different grades, a grade is assigned to the two areas that make up most of the cancer. These two grades are added to yield the Gleason score (also called the Gleason sum). The first number assigned is the grade that is the most common in the tumor. For example, if the Gleason score is written as 3 + 4 = 7, it means most of the tumor is grade 3 and less is grade 4, and they are added for a Gleason score of 7. The lowest Gleason score of a cancer found on a prostate biopsy is Gleason 6, and the highest score can be 10. Gleason score 6 and 7 cancers may be called well-differentiated/low grade and intermediate grade, respectively, and are likely to be less aggressive; that is, they tend to grow and spread slowly. Cancers with Gleason scores of 8–10 may be called poorly differentiated or high grade. These cancers are likely to grow and spread more quickly.

Chi-square test was applied to test whether there were any significant differences of proportions of stage 2 versus stage 3 (T2 versus T3) cancers in different hierarchical clusters of prostate samples (Figure 3A). Two-sided *t*-test was applied to test whether average hybridization signals of microorganisms detected were significantly different in patients with different grades or stages of cancer (Figure 3B).

Probe-capture and next-generation sequencing

The probe-capture method we used has been described previously (12,14,15,19). Briefly, the WTA products of the prostate cancer samples were pooled together into six pools for hybridization with selected biotinylated probes that had identified microbial nucleic acids in the prostate cancer samples by the PathoChip screen. The targeted sequences were then captured by Streptavidin-coated magnetic beads, and libraries were generated for NGS. Specifically, the selected probes were synthesized as 5′-biotinylated DNA oligomers (Integrated DNA Technologies, Coralville, IA), pooled together and hybridized to WTA pools of prostate cancer samples. The capture probe pool was added separately to each of the pooled WTAs of the prostate cancer samples (150 ng) in six separate reaction mixtures (Pr1–6) containing 3 M tetra-methyl ammonium chloride, 0.1% Sarkosyl, 50 mM Tris–HCl, 4 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 (1× TMAC buffer). The reaction mixtures were denatured (100°C for 10 min), followed by a hybridization step (60°C for 3 h). Streptavidin Dynabeads (Life Technologies, Carlsbad, CA) were added with continuous mixing at room temperature for 2 h, followed by three washes of the captured bead-probe-target complexes in 0.30 M NaCl plus 0.030 M sodium citrate buffer (2× SSC) and three washes with 0.1× SSC. Captured single-stranded target DNA was eluted in Tris–EDTA and used for library preparation using Nextera XT sample preparation kit (Illumina, San Diego, CA) (12,14,15,19). The five libraries were examined for quality control and submitted for NGS (Washington University Genome Technology Access Center, St. Louis) using an Illumina MiSeq instrument with paired-end 250 nucleotide reads. Adapters and low-quality fragments of raw reads were first removed using the Trim Galore software (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The processed reads were then aligned to the PathoChip metagenome, from which probes were designed, and the human genome using Genomic Short-read Nucleotide Alignment Program (GSNAP) (26) with default parameters. After alignment, we employed feature counts (27) to count how many reads aligned to each of the capture probe regions. The detailed results for these capture probes are summarized in Supplementary Table S6, available at *Carcinogenesis* Online, and visualized in Integrative Genomics Viewer (IGV) (28).

Microbial fusion detection

Microbial genomic insertion in somatic host chromosomes was determined as described previously (14,15). Prior to fusion detection, quality control of sequenced reads was performed. The Trim Galore software (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was employed for quality trimming of raw reads to remove adapters and low-quality fragments. We then used Virus-Clip (29) to identify the virus fusion sites in the human genome. Specifically, we made use of the virus genome as the primary read alignment target, and first aligned reads to the PathoChip metagenome. Some of the mapped reads may contain unaligned query sequence (soft-clipped segments). Soft-clipped reads (containing sequences of potential pathogen-integrated human loci) were then extracted from the alignment and mapped to the human genome. Utilizing this mapping information, the exact human and pathogen integration breakpoints at single-base resolution can be identified. All the integration sites were then automatically annotated with the affected human genes and their corresponding gene regions.

The host genes that supported viral genomic insertions by high sequence reads were subjected to Ingenuity Pathway Analysis (IPA) (30) that helped to combine the host genes with information obtained from the published literature to predict likely outcomes. IPA software provided a statistical significance of the association of those genes with disease outcome.

PCR validations of PathoChip detections and microbial genomic insertions

PCR primers from the conserved and/or specific regions of the microorganisms detected by the PathoChip screen were used for detection validations. The primers listed in Supplementary Table S7, available at *Carcinogenesis* Online, were self-designed from the detected probe sequences in the microbial genome. For the validation of microbial insertions, PCR primers were designed, so that the fusion junction could be amplified, one primer being designed from the microbial sequence and the other from the adjoining human gene sequence. The PCR amplification reaction mixtures for each reaction contained 200–400 ng of WTA product (pooled 50 cancer samples and pooled 15 BPH samples) and 20 pM each of forward and reverse primers (Supplementary Table S7, available at *Carcinogenesis* Online), 300 μM dNTPs and 2.5 U of LongAmp Taq DNA polymerase (NEB). DNA was denatured at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, specific annealing temperature for different set of primers (generally 3–5°C below melting temperature of the primers) for 30 s, and 65°C for 30 s. The amplicon size for each of the primer sets is shown in Supplementary Table S7, available at *Carcinogenesis* Online. The amplicons were gel extracted and sent to the Penn Genomics Analysis Core for Sanger sequencing. The electropherograms were visualized using the BioEdit program (31), and the sequences were subjected to NCBI BLAST (32) for identification.

Results

Experimental workflow for defining the microbiome signatures associated with prostate cancer

Our screening utilized both DNA and RNA extracted from 50 de-identified FFPE prostate adenocarcinoma samples and 15 de-identified FFPE prostatic tissue samples from patients with BPH, used as controls. DNA and RNA were extracted in parallel from rolls or mounted sections of each FFPE sample. The FFPE sample blocks had been stored at room temperature for 6–26 years. Nucleic acids retrieved from such blocks, even those aged >40 years, have been shown to be used for molecular analysis (20–22). The extracted RNA and DNA samples were subjected to whole-genome and transcriptome amplification and prepared for standard microarray analysis.

Using the PathoChip array, we defined the predominance of different viruses and other microbial agents associated with prostate cancer compared with controls. We generate normalized hybridization signals to all the probes for microbial agents on the chip. A *t*-test was applied to define the probes that were significantly present in cancer samples versus controls. The significance cutoff was log2 fold change > 1 and adjusted *P* value (with multiple testing corrections) < 0.05. We rank the microorganisms and viruses based on their decreasing total hybridization signal. This was determined by adding the average hybridization signal per viral family/microbial genus for each of the significantly detected probes in the prostate cancer samples. Thus, a high hybridization signal in the present study could signify a higher number of probes significantly detected for those particular family/genera and also specificity in the detection, as a strain difference would yield lower hybridization signal intensity (19).

We also determined the percent prevalence of a specific microorganism, virus and signatures of viral families among all the tumor samples (shown as dots in Figure 2). This was calculated by counting the number of cancer cases with hybridization signals greater than the average signal plus three times *SD* of dark corner or negative control probes, presented as a percentage.

It is important to note that PathoChip contains probes that are conserved across a family of viruses and probes that are specific to each virus in the family. The conserved probes can detect all members of a viral family; thus, they have the ability to detect heretofore uncharacterized members of the family. In contrast, the specific probes identify specific members of the family (19).

Finally, we validated the presence of several of the viruses and microorganisms detected by PathoChip by PCR and targeted NGS. For the targeted NGS, we used the probes that had positive results in the PathoChip screen to capture their complimentary targets from the DNA generated by whole-genome and transcriptome amplification from the cancer tissue. The captured DNA was subjected to NGS. This approach not only provided validation of identified microbial agents detected by the PathoChip screen, but also allowed detection of integrations of genomic sequences of microbial agents in the host chromosomes of prostate tumor tissue.

Microbial signatures detected in prostate cancer

The distribution of viruses and organisms by families and phyla was based on total hybridization signal and prevalence data (Figure 1). Figure 1Aa shows the distribution of general viral types that included tumorigenic, respiratory and enteric pathogens detected in the prostate tumor samples. The majority consisted of known tumorigenic viruses (41%) and viruses traditionally associated with respiratory infections (41%), whereas enteric viruses and viruses associated broadly with other diseases represented only 12 and 6%, respectively. Thirty-five percent of the viruses detected were Group I dsDNA viruses, which include the bulk of the tumor viruses (Poxviridae, Herpesviridae, Papillomaviridae and Polyomaviridae). The Group IV positive-strand ssRNA viruses made up the second largest group (23%) and included the families Picornaviridae, Coronaviridae, Flaviviridae and Astroviridae, many of which are respiratory tract viruses.

Among the bacteria genera significantly detected in the prostate tumor samples, 70% were gram negatives (Figure 1Ba), of which Proteobacteria were the most predominant phylum detected, comprising 55% of the total bacterial genera detections in the prostate tumors. The next most abundant phyla detected in the prostate tumor samples were Firmicutes (19%), followed by Actinobacteria (11%) and Bacteroides (7%).

The most prevalent fungal families detected in the prostate tumor samples were dermatophytes (31%), yeasts (15%), zygomycetes (15%) and microsporidia (12%) (Figure 1Ca). Considering the phyla (Figure 1Cb), the majority of the fungal signatures arose from the division Ascomycota (61%), 50% of which belong to the class Eurotiomycetes (Figure 1Cc).

The parasitic signatures detected in the prostate cancer samples are shown in Figure 1Da. The phylum Nematoda (36%) was most prevalent, followed by Sarcomastigophora (28%), Platyhelminthes (23%), Apicomplexa (10%) and Acanthocephala (3%). Signatures of intestinal roundworms, such as *Ancylostoma*, *Ascaris*, *Capillaria*, *Enterobius*, *Necator*, *Strongyloides* and *Trichuris*, accounted for the majority (18%) of the parasitic signatures detected, followed by signatures of other tissue roundworms, such as *Angiostrongylus*, *Contracaecum*, *Gnathostoma*, *Toxocara* and *Trichinella*, that accounted for 13% of the parasitic signature detections (Figure 1Db).

Analysis of total hybridization signal and prevalence

As described previously, we determined the total hybridization signal (cancer versus control, in blue bars) and prevalence in the sample set (in red dots) for the viral and microbial probes most abundantly detected in the cancers compared with the controls (Figure 2). These included probes of viral families, as well as specific viruses, along with specific probes for bacteria, fungi and parasites. Each was represented by the blue bar graphs in descending order of total hybridization signal per accession (for specific viruses, bacteria, fungi and parasites) or per viral family (Figure 2).

The viral families partitioned into groups representing high, medium and low total hybridization (Figure 2A, top row). The high hybridization group included (in descending order) Poxviridae, Reoviridae, Papillomaviridae and Herpesviridae. A variety of viruses were detected in the medium and low total hybridization groups; signatures of the tumorigenic Retroviridae and Polyomaviridae were detected in the medium and low total hybridization groups, respectively.

Detection of virus family was achieved by hybridization to conserved probes that detected all members of the family, for example this was the case with poxviruses. However, in the cases of papillomaviruses, retroviruses and herpesviruses, we could identify specific family members by hybridization to family member specific probes. Figure 2A, bottom row, shows the specific viruses within a family that could be detected in prostate cancer versus the non-cancerous controls.

The total hybridization signal for the papilloma viruses was only 20% compared with that for the poxviruses; however, papillomaviruses are well known to be associated with cancer. Interestingly, the high-risk HPV18 and 16 showed the highest and the third highest signals, respectively, and were detected in >80% of the samples. Numerous low-risk HPVs were detected in 60–80% of the samples.

Although total hybridization signals for the retroviruses family were only in the medium hybridization range, we examine individual retroviral signatures due to the significance of retroviruses in cancer (Figure 2A, bottom row). Lentivirus showed the highest signal among the retrovirus signatures followed by foamy virus. Detected at much lower signals were mouse mammary tumor virus (MMTV), Moloney murine leukemia virus (MMLV) and human T-lymphotropic virus (HTLV). Among the human herpesvirus (HHV), HCMV was the most represented

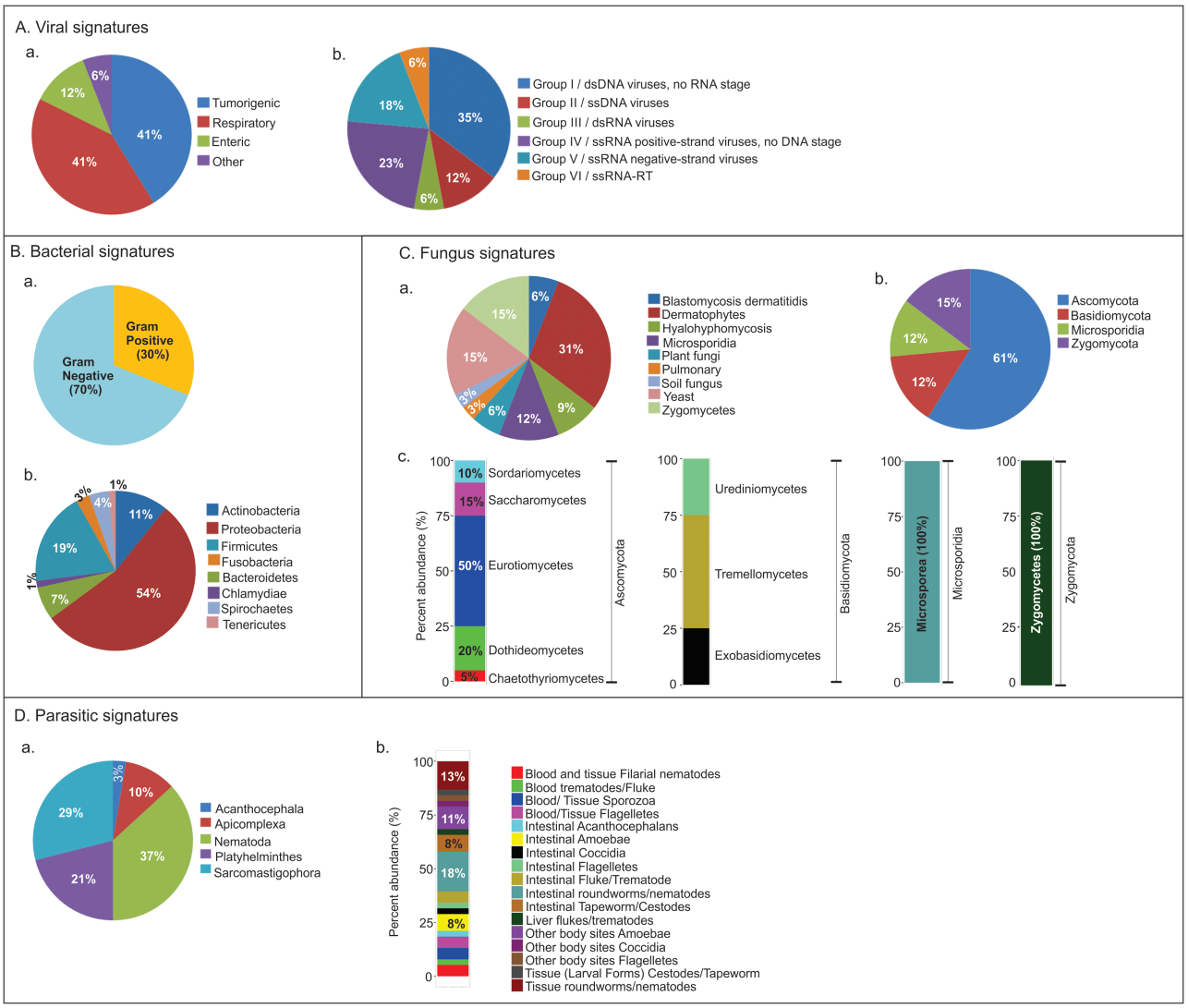

Figure 1. Distribution of microbial signatures significantly associated with prostate cancer samples compared with controls. (**A**) Proportion of different viral signatures detected significantly in the prostate cancer samples are represented as pie charts, showing categories of different viral types (a) and groups (b) of different viruses detected. (**B**) Proportion of different bacterial signatures detected significantly in the prostate cancer samples are represented as pie charts, showing the percentage of different groups and phyla of bacteria detected significantly in prostate cancer samples. (**C**) Proportion of different fungal signatures detected significantly in the prostate cancer samples, showing the percentage of different types (a), phyla (b) represented as pie charts, as well as class (c) of fungi detected significantly in prostate cancer samples. (**D**) Proportion of different parasitic signatures detected significantly in the prostate cancer samples, showing the percentage of different phyla (a) as pie chart, and types (b) of parasites detected significantly in prostate cancer samples.

by total hybridization signal followed by Kaposi sarcoma associated herpesvirus (KSHV). Other specific herpesvirus signatures detected were that of HHV6B, HHV3 and HHV7.

We similarly analyzed the bacteria, fungi and parasites detected specifically in prostate tumors (Figure 2B–D). In each case, they are partitioned into high, medium and low total hybridization groups, where the specific microbial genera were seen in at least 30% of the samples (prevalence in red dots).

Of the 13 bacterial genera detected with high hybridization signals, 9 genera are Proteobacteria, including *Rickettsia* (detected in 80% of the samples), *Bordetella*, *Sphingomonas*, *Bartonella*, *Helicobacter*, *Salmonella*, *Aeromonas*, *Brevundimonas* and *Shigella*. The next predominant phyla detected were gram-positive Firmicutes (18%; Figure 1Bb) including *Bacillus*, *Lactobacillus*, *Enterococcus*, *Clostridium*, *Pediococcus*, *Streptococcus*, *Peptoniphilus*, *Listeria*, *Aerococcus*, *Lactococcus*, *Staphylococcus*, *Abiotrophia* and *Geobacillus*. In most cases, the Firmicutes were detected at medium to low hybridization signal intensity, the exception being *Bacillus*, which was detected in the high hybridization signal intensity group (Figure 2B). *Mycobacterium*, a gram-positive Actinobacteria, was detected with high hybridization signal intensity in 72% of the prostate cancer samples.

The fungi that had high, medium and low hybridization signals in the tumor samples are represented in Figure 2C. The high hybridization signal group contained *Alternaria*, followed by *Malassezia*, *Candida* and *Cladosporium* in >82% of the cancer samples (Figure 2C). Among the other signatures that were detected in the high hybridization group were *Trichosporon*, *Cladophialophora*, *Rhodotorula*, *Geotrichum*, *Fusarium*, *Mucor*, microsporidia like *Nosema* and *Pleistophora* (Figure 2C). Figure 2C also shows that a number of other fungi signatures were detected in the medium and low hybridization signal groups in varying percentages of the samples. Signatures of other genera of yeasts, like *Cryptococcus*, *Buckleyzyma* and *Issatchenkia*, were

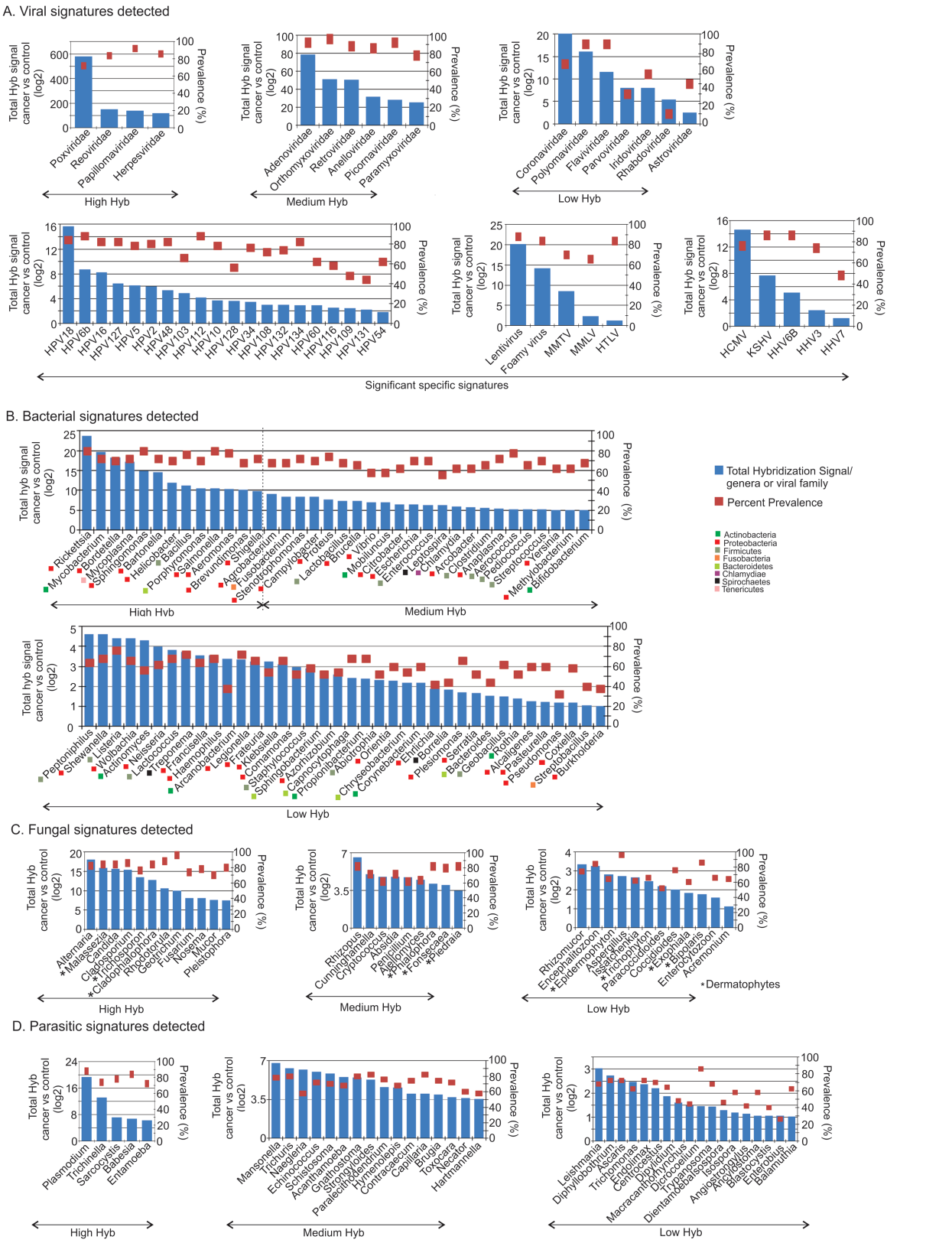

Figure 2. Microbial signatures detected significantly in the prostate cancer samples versus the controls by PathoChip screen. (**A**) Signatures of viral families and predominant specific viral signatures detected in prostate cancer are shown, with the total hybridization signal (sum of hybridization signals for all the detected viruses in the family, or all the significantly detected probes for a specific virus) for each viral families or for each specific virus represented according to descending order as a bar graph and the prevalence of the same as dots. The known tumorigenic viral families detected are indicated. HPV, human papilloma virus, MMTV, mouse mammary tumor virus; MMLV, Moloney murine leukemia virus; RSV, Rous sarcoma virus, ACV, avian carcinoma virus; HTLV, human T-lymphotropic virus; KSHV, Kaposi sarcoma-associated herpesvirus or human herpes virus 8, HHV, human herpes virus. Representation of detected bacterial (**B**), fungal (**C**) and parasitic (**D**) signatures in prostate cancer as bar graphs, with the total hybridization signal for all the significantly detected probes per genera shown in descending order and prevalence of the same as dots.

also detected, but with medium and low hybridization signals in 62% of cancer cases (Figure 2C). Similarly, another signature of microsporidia, *Fonsecaea*, was detected in >80% of the cancer samples, but with medium hybridization signal (Figure 2C).

The total hybridization signals of signatures for each of the genera of parasites significantly detected in the prostate cancers are represented in Figure 2D. Signatures of blood/tissue Sporozoa, *Plasmodium*, was detected with the highest hybridization signal intensity in 88% of the cancer samples screened (Figure 2D). This was followed by *Trichinella*, *Sarcocystis*, *Babesia* and *Entamoeba*, all of which were detected in >70% of the samples screened (Figure 2D). Signatures of a number of parasites were also detected in the medium and low hybridization signal groups in varying percentages of the samples (20–85%) (Figure 2D). In particular, signatures of *Trichuris*, *Naegleria*, *Echinococcus*, *Schistosoma*, *Strongyloides*, *Hymenolepis*, *Contracaecum* and *Toxocara* in the medium hybridization signal group, and *Enterobius* in the low hybridization signal group, are notable.

For the abovementioned significant detections in the prostate cancers, the median hybridization signal that represents the mid-value for hybridization signals of all the detected probes per accession/viral families is shown in Supplementary Figure S1A and Table S2A, available at *Carcinogenesis* Online, and the percentage of significant probes per accession that were more highly detected in the cancers compared with the controls are shown in Supplementary Table S2C, available at *Carcinogenesis* Online. Among the different microbial signatures detected, viral signatures showed the highest prevalence (Supplementary Figure S1B, available at *Carcinogenesis* Online).

Although we detected many microbial signatures in the prostate controls (Supplementary Figure S2 and Table S3, available at *Carcinogenesis* Online), only a few microbial signatures were detected significantly higher in the controls compared with the cancers (Supplementary Figure S1C, available at *Carcinogenesis* Online). They included a few conserved signatures of Retroviridae, Poxviridae, Reoviridae and Herpesviridae, along with specific probes of bacteria *Chlamydia*, *Pseudomonas*, *Burkholderia*, *Campylobacter* and parasite *Babesia* (Supplementary Figure S1C, available at *Carcinogenesis* Online). However, a few probes of *Helicobacter* were detected (hybridization signal > 1) sporadically in the controls (Supplementary Table S3, available at *Carcinogenesis* Online), but the average hybridization signal of detection of the *Helicobacter* probes was significantly lower in the controls compared with the cancers (Supplementary Table S3, available at *Carcinogenesis* Online); thus, it was not reported as a part of the prostate control/BPH-associated microbiome.

However, screening tissue-free paraffin-detected bacterial (*Propionibacterium*, *Sphingobacterium*, *Chryseobacterium* and *Capnocytophaga*) and fungal (*Alternaria* and *Malassezia*) signals (Supplementary Table S2C, available at *Carcinogenesis* Online), which suggest that they were included in the sample processing. There was no significant detection of microorganism probes detected in a non-template negative control screen (water instead of DNA/RNA; data not shown). These results indicate that probes detected in the tissue-free paraffin sample may be a consequence of the embedding process. Thus, these signatures were not included as a part of the prostate cancer-specific microbiome.

Clustering of prostate cancer samples based on the detection of microbiome signature patterns

Hierarchical clustering based on microbial signature detection patterns were represented in two different ways (Figure 3, left and right panels). Figure 3 (left panel) shows the heat map of the hierarchical clustering where prostate cancer samples with similar hybridization signal for microbiome signatures cluster together. In contrast, Figure 3 (right panel) shows the best numbers of such clusters that can be formed from the data set using the same algorithm. Hierarchical analysis concluded that the 50 prostate cancer samples grouped into two main clusters (cluster 1 and 2), and samples in cluster 2 were again broadly subdivided into two distinct subgroups (subcluster 2a and 2b). There was no significant change in clustering analysis after we eliminated the signals detected from the empty paraffin screen from the analysis (Supplementary Figure S3, available at *Carcinogenesis* Online).

The microbial signatures that were common to all the clusters/subcluster (ANOVA test, $P < 0.05$) were bacterial signatures of *Capnocytophaga*, fungal signatures of *Bipolaris* and parasitic signatures of *Dipylidium* and *Angiostrongylus* (Supplementary Table S4, available at *Carcinogenesis* Online, ANOVA test).

Prostate cancer cases in each cluster/subcluster differed based on the detection of certain significant microbial signatures (two-tail *t*-test between individual clusters versus the rest, adjusted $P < 0.05$) (Supplementary Table S4, available at *Carcinogenesis* Online, cluster 1 versus rest, subcluster 2a versus rest, subcluster 2b versus rest). Prostate cancer samples in cluster 1 had significant higher detection for most of the microbial signatures detected when compared with cluster 2 samples (Supplementary Table S4, available at *Carcinogenesis* Online, cluster 1 versus rest), and in fact, cancer samples in cluster 1 showed >2-fold (log2 fold change > 1) higher detections for certain bacterial (*Actinomyces*, *Aerococcus*, *Alcaligenes*, *Arcanobacterium*, *Geobacillus*, *Klebsiella*, *Plesiomonas*, *Propionibacterium*, *Rothia*, *Serratia*, *Sphingobacterium*, *Staphylococcus* and *Enterobius*) and parasitic (*Dicrocoelium* and *Dientamoeba*) signatures. Among the cluster 2 cancer samples, subcluster 2a had higher detections of all the microbial signatures compared with cluster 2b samples (Supplementary Table S4, available at *Carcinogenesis* Online, subcluster 2b versus 2a). In fact, fungal signatures of *Cladophialophora*, *Fonsecaea*, *Phialophora*, *Piedraia* and parasitic signatures of *Prosthodendrium*, *Centrocestus* and *Trichuris* were significantly detected higher in subcluster 2a than the rest (cluster 1 + subcluster 2b) of the cancer samples (Supplementary Table S4, available at *Carcinogenesis* Online, subcluster 2a versus rest). Although prostate cancer samples that grouped in subcluster 2b showed significant lower detection for most of the microbial signatures (Supplementary Table S4, available at *Carcinogenesis* Online, subcluster 2b versus rest), the average hybridization signal intensity for certain bacterial (*Rothia*, *Geobacillus*, *Actinomyces*, *Arcanobacterium*, *Brevundimonas*, *Peptoniphilus*, *Klebsiella*, *Frateuria*, *Mobiluncus*, *Serratia*, *Francisella*, *Burkholderia*, *Vibrio*, *Aerococcus*), fungal (*Piedraia*) and parasitic (*Macracanthorhynchus*) signatures in subcluster 2b was found to be >2-fold lower (log2 fold change < –1) compared with the rest of the cancer samples. Those signatures were detected with low hybridization signal intensity in only a few of the subcluster 2b samples. A small number of samples in subcluster 2b had very high detection of the parasitic signatures of *Dipylidium* (Supplementary Table S4, available at *Carcinogenesis* Online, subcluster 2b versus 2a, log fold change for *Dipylidium* = 0.01). Heatmap for these differentially detected signatures between clusters is shown in Supplementary Figure S4, available at *Carcinogenesis* Online.

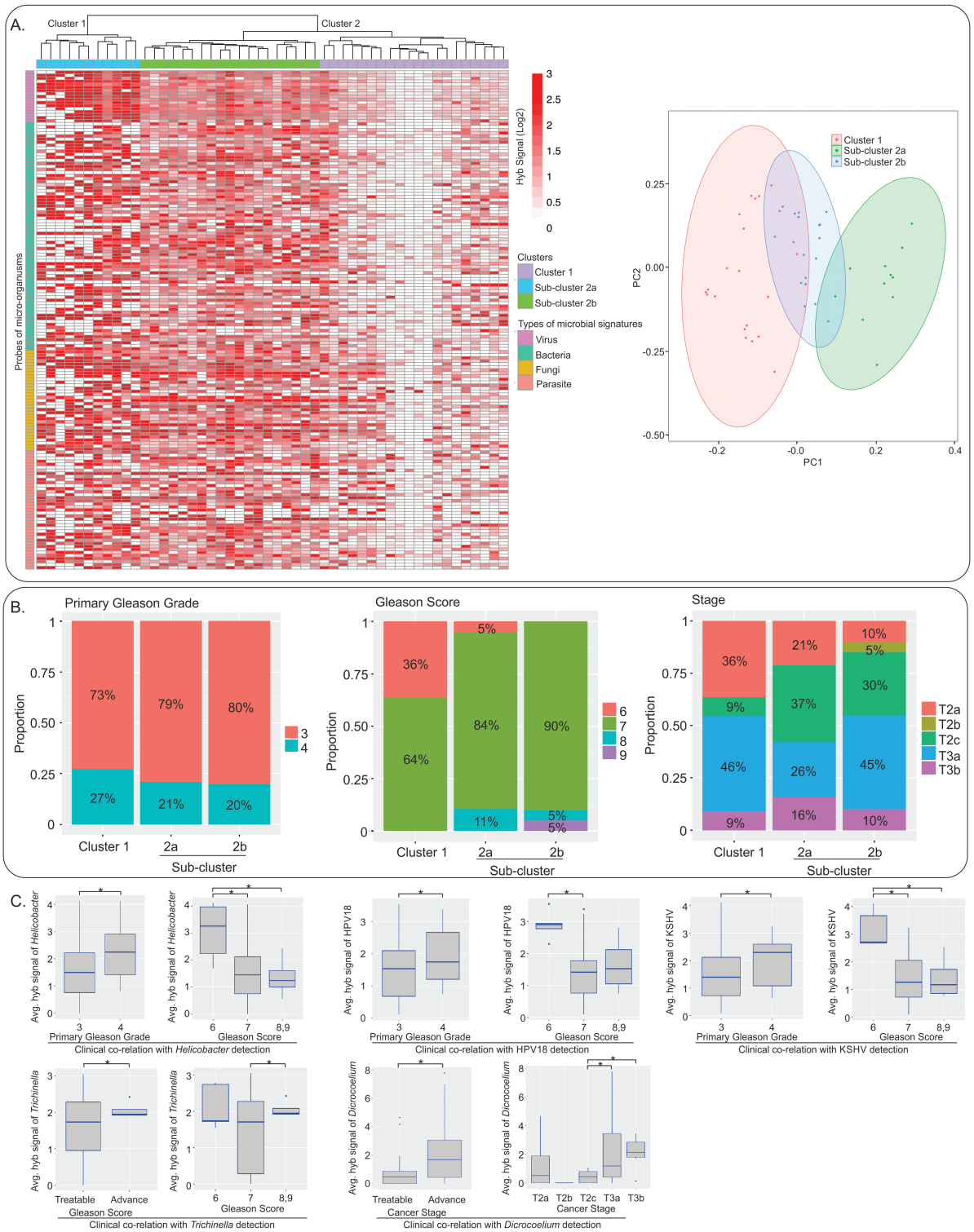

Figure 3. Hierarchical clustering of prostate cancer samples along with clinical correlations. (**A**) Hierarchical clustering based on microbial signature detection pattern in 50 prostate cancer samples, represented as a heat map (left). Clustering was performed by R program using Euclidean distance, complete linkage and non-adjusted values. Clustering of the prostate cancer samples using NBClust software [Calinski and Harabasz index, Euclidean distance, complete linkage] (right). (**B**) Proportions of prostate cancer patients in each of the hierarchical clusters having different grades, scores and stages of prostate cancers. (**C**) Specific microbial signature detections in different grades, scores and stages of prostate cancer patients. Only the significant ($P < 0.05$) differences between different grades, scores and stages are shown with an 'asterisk' (*). The *P* values are as follows: *Helicobacter* detection in cancer stages 6 versus 7 ($P = 0.031$), between 6 versus 8,9 ($P = 0.032$) and between primary Gleason score 3 versus 4 ($P = 0.084$); HPV18 detection in cancer stages 6 versus 7 ($P = 0.00023$) and between primary Gleason grade 3 versus 4 ($P = 0.0017$); KSHV detection in cancer stages 6 versus 7 ($P = 0.0018$), between 6 versus 8,9 ($P = 0.0139$) and between primary Gleason score 3 versus 4 ($P = 0.00006$); *Trichinella* detection in cancer stages 7 versus 8,9 ($P = 0.02$) and between treatable versus advance ($P = 0.032$); *Dicrocoelium* detection in cancer stages T3a versus T2c ($P = 0.006$), T3b versus T2c ($P = 0.017$) and between treatable versus advance ($P = 0.009$).

Correlation of clustered microbiome signatures and cancer grade and stage

Using the limited clinical data provided (Supplementary Table S1, available at *Carcinogenesis* Online), we were able to document trends comparing prostate cancers in specific hierarchical clusters with the reported Gleason grades, Gleason scores and the reported stages of the cancer.

The number of prostate cancer cases in each of the hierarchical clusters with different scores and stages is shown in Supplementary Table S5, available at *Carcinogenesis* Online. Figure 3B (left panel) compares primary Gleason grades 3 and 4 where these is no significant difference in the distribution of the primary Gleason grade type between the clusters.

Figure 3B (middle panel) shows the distribution of Gleason scores of the samples in each cluster. Although a majority of the samples in each cluster were scored 7, cluster 1 samples had the lowest percentage scoring 7 (moderately differentiated or intermediate grade), with all the remaining scored 6 (well-differentiated or low grade). Hence, cluster 1 tends toward lower Gleason scores. In contrast, clusters 2a and 2b tended toward greater numbers of grade 7 and higher scores of 8–9 (poorly differentiated tumor tissues with high risk of advanced cancer). Thus, higher hybridization signals for the microbiome signatures, as in cluster 1, correlated with better Gleason scores. Correspondingly, cluster 2 samples with moderate (cluster 2a) to low (cluster 2b) hybridization signal for the microbiome signatures had higher Gleason scores.

Figure 3B (right panel) shows the distribution of the diagnosis of stage 2 (T2a, T2b and T2c) or 3 (T3a and T3b) prostate cancers in the different hierarchical clusters. Here we noted that cluster 1 and subcluster 2b had almost similar distribution for stage 2 and 3 cancers.

However, the distribution between T2a, T2b and T2c for stage 2 was quite different with the majority being T2a in cluster 1 patient groups and T2c in subcluster 2a and 2b groups. Samples in subcluster 2a had a slightly higher number of stage 2 but fewer stage 3, although this cluster did have the most cases in the advanced stage T3b.

Further analysis shows that certain microbial signatures are significantly higher in different grades and stages of cancer (Figure 3C; Supplementary Table S5, available at *Carcinogenesis* Online). Average hybridization signal of signatures of Astroviridae, *Borrelia*, *Candida*, *Capillaria*, *Entamoeba*, *Enterobius*, *Histoplasma*, *Legionella*, *Mansonella*, *Porphyromonas*, *Shigella* and *Streptobacillus* were significantly higher in prostate cancer with lower Gleason scores (scores 6 and 7) (Supplementary Table S5, available at *Carcinogenesis* Online). Conversely, the signature of *Trichinella* was significantly higher in prostate tumors with higher Gleason score (scores 8 and 9) (Figure 3C; Supplementary Table S5, available at *Carcinogenesis* Online). In addition, we found the signatures of *Helicobacter*, HPV18, KSHV and polyomaviridae family members to be higher in prostate cancer with lower Gleason score (score 6) than the ones higher than 6 (Figure 3C). Examining stages, we found that the signature of *Dicrocoelium* were significantly higher in stage 3 (T3) prostate cases than stage 2 (T2) (Figure 3C; Supplementary Table S5, available at *Carcinogenesis* Online).

These comparisons suggest that there is a possibility that certain microbial signatures in prostate cancers can predict clinical diagnosis and potential outcomes of disease. Thus, a specific microbial signature in prostate has potential prognostic/diagnostic value.

Validation of PathoChip screen results

Array probes representing Adenoviridae, Papillomaviridae and Herpesviridae that were positive in the PathoChip screen were used to capture complimentary sequences from the whole-genome and transcriptome-amplified pooled prostate cancer samples. The captured DNA was then subjected to NGS. The resulting sequences were aligned with the PathoChip metagenome. This metagenome comprises the concatenated genomic sequences of all the viruses and microorganisms from which the probes on the PathoChip were generated. This analysis showed alignment with the capture probe locations on the metagenome, which validates the accuracy of the probes for the specific virus. Figure 4A, Supplementary Figure S5 and Table S6, available at *Carcinogenesis* Online, show the sequence alignments in six separate capture reactions (Pr1–6) to the HPV18 and HHV8 region of the PathoChip metagenome. The sequences marked with a circle (o) in Figure 4A are the sequence reads that aligned to the capture probe locations. There were also sequences that aligned with other probes in the same accession that were not used as the capture probes (marked with an asterisk '*'). This may result from pull-down of larger fragments of microbial genomic sequences by the capture probes. Supplementary Table S6, available at *Carcinogenesis* Online, shows the number of reads that aligned to the capture probe locations (inProbe) and also outside (outProbe) of it for the accessions in each capture reactions.

Validation of other microbial signatures was carried out by PCR and Sanger sequencing (Figure 4B; Supplementary Figure S6, available at *Carcinogenesis* Online). The primers used for validation of the PathoChip screen results by PCR are in Supplementary Table S7, available at *Carcinogenesis* Online, and are designed from the region of the microbial genome that tested positive by PathoChip screen. These primers were used to directly amplify the microbial signatures from the pooled whole-genome and transcriptome-amplified product (WTA) of the 50 prostate cancer samples, which were being used for the hybridization step in the PathoChip screen.

Detection of viral genomic insertions in host chromosomes of prostate cancer samples

Some of the sequences obtained from probe-capture NGS only partially aligned with the PathoChip metagenome. The flanking non-aligning regions were found to align with human sequences instead, thus suggesting sites of viral and microbial DNA integration in human chromosomes. Using Virus-Clip (29) (see Materials and methods) to detect such integrations, we detected multiple integration sites in the host chromosomes for a number of herpesviruses and papillomaviruses (Figure 5, Supplementary Table S8, available at *Carcinogenesis* Online). The prevalence of particular viral genomic integrations in the host chromosomes is shown in Figure 5A, right panel. The highest number of genomic integration sites was detected for HPV18 (33), followed by KSHV (34), HPV2, HPV108, a herpesvirus with a signature similar to Equid herpesvirus 9 and herpesvirus 2 (Figure 5A). Figure 5B shows the sites of integration within the viral genomes. These data suggest that some of the viruses have preferential regions for integration, for example *E1* in HPV18 (61%), *L2* in HPV108 (53%) and *UL42* of herpesvirus 2 (60%), and are discussed below. The integration sites in the host chromosomes for these viruses are represented in a circos plot (Figure 5D, left panel) and karyotype plot (Figure 5D, right panel), and are also mentioned in Supplementary Table S8, available at *Carcinogenesis* Online. Furthermore, we were analyzing results from a pool of

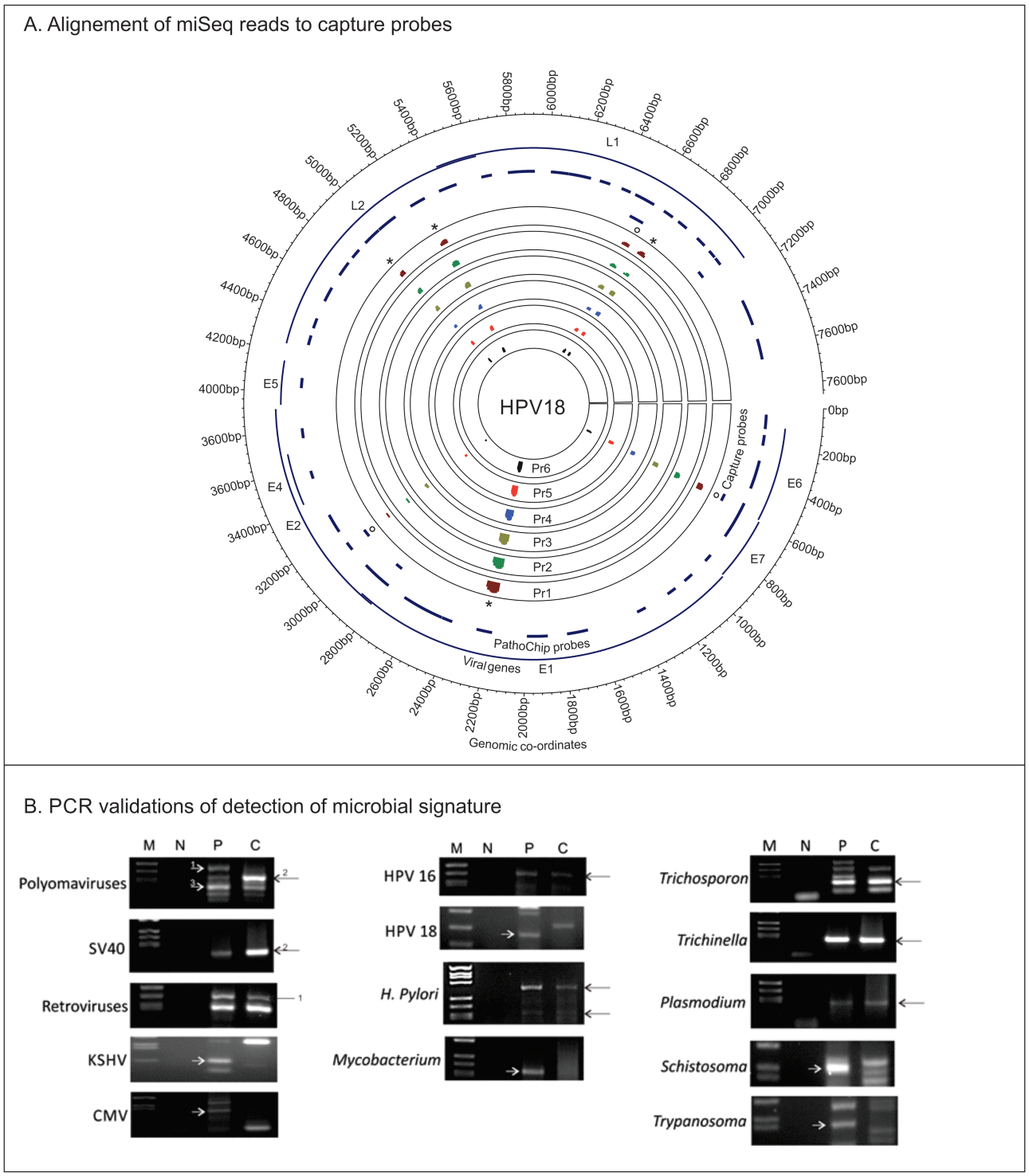

Figure 4. Validations of PathoChip screen results. (**A**) Probe-capture sequencing validation of PathoChip detection of HPV18. Targeted MiSeq reads aligned to capture probe locations of HPV18 genome, represented as a circos plot. Probe-capture sequencing alignment is shown for individual capture pools (Capture 1–6 or Pr1–6). The whole-genome amplified DNA plus cDNA of the prostate cancer samples were hybridized to a set of biotinylated probes, then captured by streptavidin beads, and used for tagmentation, library preparation and deep sequencing with paired-end 250 nucleotide reads. The miseq reads from individual capture when aligned with the metagenome of PathoChip (PathoChip probes) was found to cluster mostly at the capture probe regions. The genomic co-ordinates with viral genes are mentioned in the figure for HPV18. (**B**) PCR validations of PathoChip screen results. Using the detection primers mentioned in Supplementary Table S6, available at *Carcinogenesis* Online, PCR was carried out to validate the PathoChip screen detection of the microbial signatures in the prostate cancer samples. Figure 4B shows the gel pictures of EtBr-stained amplicons run on 2% agarose gel, where M is DNA ladder of RsaI digested φX/174, NTC is non-template control, P is pooled WTA product of prostate cancer samples and C is pooled WTA product of prostate control samples.

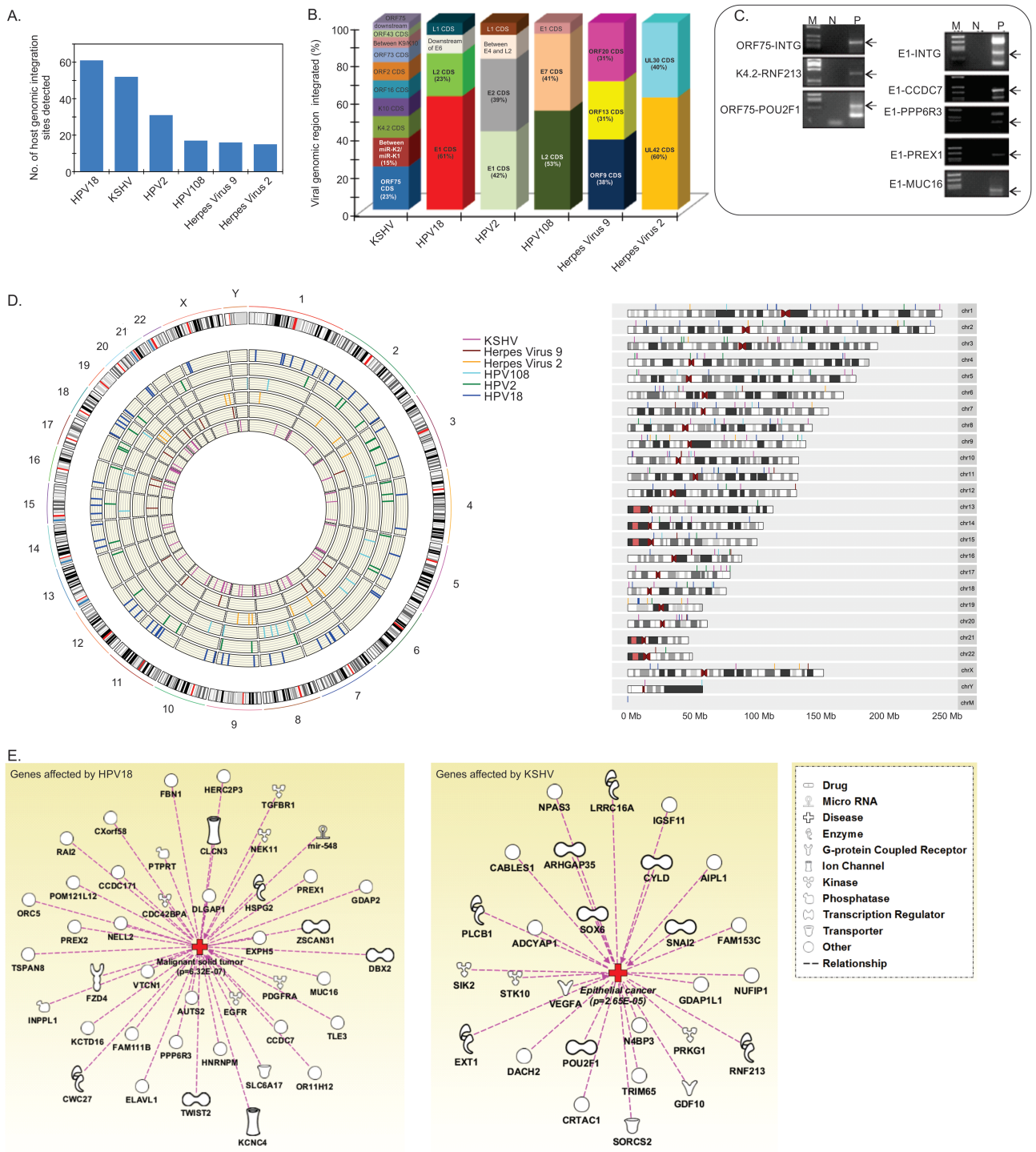

Figure 5. Microbial genomic integrations in the host chromosome. (**A**) Bar graph showing the number of host genomic integration sites detected in the study for HHV and HPV genomic insertions. (**B**) Bar graph showing the percentage of different viral genomic regions of HHV and HPV, which were found to be integrated in the host somatic chromosomes. (**C**) PCR validation of KSHV and HPV18 genomic fragment insertions in the host chromosomes, using primers for insertion validations listed in Supplementary Table S6, available at *Carcinogenesis* Online. Primers were designed such to amplify the host-viral genomic junction. KSHV *ORF75* sequence insertions at an intergenic (INTG) and intronic region of *POU2F1* gene, KSHV *K4.2* sequence insertion at the exonic region of *RNF213* gene, HPV18 *E1* insertion at an intergenic region (INTG), at the intronic region of *CCDC7* gene, downstream of *PPP6R3* gene, at the intronic region of *PREX1* and *MUC16* genes are validated by PCR. (**D**) Circos plot highlighting fusion events with ≥20 reads support for the detected viral insertions into individual human chromosomes are shown. Different HHV and HPV insertions sites on individual human chromosomes are represented by differentially colored lines on each of the concentric circles on the plot. (**E**) Karyogram plot of the HHV and HPV viral insertion sites in human chromosomes, cutoff reads ≥ 20, represented by differentially colored lines for each type of viral insertions. (**F**) Significant ($P < 005$) association of most of the host genes affected by HPV18 and KSHV/HHV8 genomic integrations to cancer as analyzed by IPA program.

prostate

cancer patient samples. Thus, integrations detected are the total from a heterogeneous population of cells. The integrations were seen at the exonic, intronic, intergenic, UTR, upstream and downstream of many genes (Supplementary Table S8, available at *Carcinogenesis* Online). It is possible that viral genomic insertion into these genes could affect their expression. Using IPA (QIAGEN Redwood City, www.qiagen.com/ingenuity), we found that many of the genes are associated with tumorigenesis (Figure 5E). To validate the insertions we randomly selected several of the detected KSHV (HHV8) and HPV18 genomic insertions and used PCR with primers designed to amplify the host-viral junction region (Supplementary Table S7, available at *Carcinogenesis* Online), followed by Sanger sequencing of the amplicons (Figure 5C; Supplementary Figure S7, available at *Carcinogenesis* Online).

HPV18 genomic integrations detected in the host chromosomes of prostate cancer

Of the different HPV18 genomic regions found integrated in the prostate cancer chromosomes, there were two prevalent hotspots for insertion. Sixty-one percent of the insertions occurred within the *E1* coding sequence region at genomic location 2238 in the accession NC_001357.1. The second hotspot, representing 23% of the insertions, was within the L2 coding region at genomic location 4946 (Figure 5B; Supplementary Table S8, available at *Carcinogenesis* Online).

The majority of the HPV18 genomic integrations were detected in intergenic regions (46%), followed by intronic regions (38%) (Supplementary Table S8, available at *Carcinogenesis* Online). Integrations were detected upstream and downstream of genes, and one insertion was detected within the 3′UTR of *ELAVL1* gene. HPV18 *L2* coding sequences were found inserted upstream of the *FAM111B* and *KCNC4* genes (Supplementary Table S8, available at *Carcinogenesis* Online) both of which were found to be associated with malignant solid tumor formation (Figure 5E, left panel). These insertions may affect gene expressions. HPV18 insertions were detected within intronic regions of 20 genes (Supplementary Table S8, available at *Carcinogenesis* Online). Eighteen of these were found to be associated with malignant solid tumor formation (Figure 5E, left panel). These included *CCDC7, PREX1, EXPH5, MUC16, HSPG2, NEK11, HNRNPM, CWC27, CDC42BPA, PREX2, AUTS2, VTCN1, ZSCAN31, PTPRT, CLCN3, DLGAP1, CCDC171* and *FBN1*. It is possible that these intronic integrations may affect gene expression due to alterations in splicing during transcription (35).

Other HPV genomic integrations detected in the host chromosomes of prostate cancer

Coding sequences for both HPV2 (NC_001352.1) *E1* (genomic co-ordinate 2132 and 2696) and *E2* (genomic co-ordinate 2696 and 3550) appear to be highly susceptible to genomic integrations (Figure 5B; Supplementary Table S8, available at *Carcinogenesis* Online). Of the 31 HPV2 integration sites detected in prostate cancer chromosomes, 12 are at intronic regions, 16 are within intergenic regions, 2 are downstream of genes and 1 is within the 3′UTR of the *IL18BP* gene (Supplementary Table S8, available at *Carcinogenesis* Online). Ten of the 12 genes with intronic HPV2 insertions were significantly (P = 3.12E-04) associated with adenocarcinoma by IPA (Supplementary Figure S8A, available at *Carcinogenesis* Online). Viral genomic integrations in intergenic regions can affect neighboring gene expression if within 100 kb of the genes (36,37). Seven of nine such genes, with HPV2 insertions, are within 100 kb and are also found to be significantly (1.72E-02) associated with adenocarcinoma by IPA (Supplementary Figure S8B, available at *Carcinogenesis* Online).

In case of HPV108 genomic integrations in the host chromosomes of prostate cancer cells, the majority were in the region within the *L2* coding sequence (genomic co-ordinate 4771 of NC_012213.1) (53%), followed by *E7* coding sequence (genomic co-ordinate 944 of NC_012213.1) (41%) (Figure 5B, available at *Carcinogenesis* Online). Integrations were seen in both intronic and intergenic regions. The majority of affected genes were found to be significantly (P = 5.25E-03) associated with epithelial cancers (Supplementary Figure S8C, available at *Carcinogenesis* Online).

KSHV genomic integrations detected in host chromosomes of prostate cancer

Among the KSHV genomic regions found to be integrated, the highest (23%) integrations involved the *ORF75* coding region at the co-ordinate 130802 of accession NC_009333.1. This was followed by the region between miR-K2 and miR-K1 at the co-ordinate 122116 (Figure 5B; Supplementary Table S8, available at *Carcinogenesis* Online).

The majority of the KSHV insertions were found within intergenic (50%) and intronic regions (42%) (Supplementary Table S8, available at *Carcinogenesis* Online). Among the intergenic integrations, 54% were seen within 100 kb upstream or downstream of genes (see below). The rest were in regions upstream of the *GDF10* gene, downstream of the *CYLD* gene and within an exonic region of the *RNF213* gene, all of which are correlated with epithelial tumorigenesis (Supplementary Table S8, available at *Carcinogenesis* Online).

KSHV intergenic insertions within 100 kb of coding regions included 40 kb upstream of *HNF4G*, 12 kb upstream of *CABLES1* gene, 49 kb upstream of *LINC01512* gene, 47 kb upstream of *LOC728084*, 55 kb upstream of *FLJ36777* gene, 34 kb upstream of *SORCS2* gene, 80 kb upstream of *MIR3910-1* gene, 29 kb downstream of *AIPL1*, 76 kb downstream of ncRNA *LOC101927964*, 42 kb downstream of ncRNA *LINC00702*, 28 kb downstream of *ADCYAP1*, 23 kb downstream of *TMEM241*, 55 kb downstream of *VEGFA*, 32 kb downstream of *FAM153C*, 32 kb downstream of *N4BP3*, 89 kb downstream of ncRNA *LINC01247*, 39 kb downstream of *SNAI2*, 23 kb downstream of *LOC100506499*, 20 kb downstream of ncRNA *C14orf183/LINC01599* and 60 kb downstream of *SOX6*. The majority of these genes were found to be significantly correlated (P = 2.65E-05) with the development of epithelial cancers as determined by IPA (Figure 5E, right panel).

Other herpesvirus genomic integrations detected in the host chromosomes of prostate cancer

Integration sites within the genomes of other herpes viruses included conserved sequences of the cercopithecine herpesvirus 2 (NC_006560.1) *UL42* gene (genomic co-ordinate 92081), which encodes a DNA polymerase processivity subunit, and the coding sequence of the *UL30* gene (genomic co-ordinate 65234), which encodes DNA polymerase catalytic subunit. In addition, coding sequences of *ORF9* (genomic co-ordinate 110214), *ORF13* (genomic co-ordinate 15424) and *ORF20* (genomic co-ordinate 110214) of Equid herpesvirus 9 (NC_011644.1) (Supplementary Table S8, available at *Carcinogenesis* Online) were also found to be integrated in the host chromosome of prostate cancer.

Helicobacter pylori cagA gene integrations were detected in host chromosomes of prostate cancer

Helicobacter pylori is a known human pathogen associated with gastric carcinoma and was previously detected in prostate

cancer (38,39). Therefore, we further examined detection of this bacterial agent in prostate tumors. To validate *H.pylori* detection in the prostate cancer samples, we used primers specific for the *cagA* gene of *H.pylori* (Supplementary Table S7, available at *Carcinogenesis* Online). We observed multiple bands in the prostate cancer samples, even under stringent PCR conditions (Figure 6). We thus sequenced the three distinct bands and subjected them to BLAST against the *H.pylori* genome. The sequence included not only *cagA* sequences, but also sequence matching the human genome. This suggested that *H.pylori* sequences were inserted in the prostate tumor cell genome. The BLAST results for the upper most band (1 in Figure 6) revealed that *cagA* sequences were inserted at the lncRNA *LOC105376839* gene on chr17 (17q21.31). The BLAST for the middle band (2 in Figure 6) showed partial *cagA* gene sequences fused with protein phosphatase 1 regulatory subunit 9A (*PPP1R9A*) gene, also called neurabin I, on chr7 (7q21.3). Finally, the sequence of the lower band (3 in Figure 6) showed integration of *cagA* sequences in neural cell adhesion molecule 1 (*NCAM1*) gene on chr11 (11q23.2).

Discussions

In previous studies, we defined the microbial signatures associated with different breast cancer types (12,13), oral cancer and ovarian cancer (14,15). These studies showed that the tumor microbiome is quite diverse compared with normal tissue and that different tumor types have distinguishing microbiome signatures. The microbiome in prostate tumors was detected at significantly higher levels than in control tissue; however, the levels of viruses and microorganisms in the prostate tumor microbiome were still relatively low. Thus, it is unlikely that we are detecting a meaningful replicative infection. We suggest that although the presence of the tumor microbiome may affect the course of the cancer, it is likely that the tumor microenvironment provides a specialized niche in which viruses and microorganisms can persist more readily than in normal tissue. In the present study, we examined the microbiome of prostate cancer compared with prostate tissue from BPH patients and reported a diverse and distinct prostate tumor microbiome compared with that of the controls.

Our study is not the first to suggest the presence of viruses and bacteria in the prostate. Several studies have documented that viral and bacterial infection of the prostate are risk factors associated with prostate cancer development (1,7,9,18,40–43).

In this regard, our data suggest a prominent representation of tumor viruses in the prostate cancer samples compared with the BPH controls. It is noteworthy that many of the viral families were detected by PathoChip probes that represent conserved regions found in all members of the virus family. Probes for specific family members were the least represented, possibly suggesting significant strain variation among individual family members. Alternatively, the higher detection by conserved probes may suggest heretofore uncharacterized members of the virus family. Our data neither support nor deny that these viruses have any direct consequence in prostate oncogenesis (44). Previous studies (7,9,18,40,42,45–47) support our findings of viral signatures in prostate cancers, these include the detection of HPV [including HPV18 (11)], HCMV, EBV, JCV and BKV. Also, our detection of the oncogenic papillomavirus HPV18 in prostate cancer has been reported (11). However, reports of the association of viruses with prostate cancer have been controversial, for example, the association of the endogenous retrovirus, xenotropic murine leukemia-related virus in familial prostate cancer patients (43,48). In this regard, we did not detect xenotropic murine leukemia-related virus, but did detect signatures of other endogenous retroviruses, such as the alpharetrovirus RSV, betaretrovirus MMTV and gammaretrovirus MMLV. In agreement, a previous study reported that MMTV-like virus DNA was found in 36% of prostate cancers tested (45). Overall, past studies support our finding of a diverse virome in prostate cancer.

Considering the bacterial microbiome, several previous studies have reported an abundance of Proteobacteria associated with dysbiosis-related diseases including cancer (49–51). A recent study using ultra-deep pyrosequencing showed the dominance of Actinobacteria in cancerous, pre-cancerous and non-cancerous prostate tissues, *Propionibacterium* being the most abundant, followed by *Corynebacterium* (17). Most of the bacterial genera detected in that study were also detected in the present study. An association of *Chlamydia trachomatis* and *P.acnes* has correlated with increased risk for prostate cancer development due to their pro-inflammatory host responses (52,53). We detected signatures of both, with medium and low hybridization signal in at least 85% of the prostate cancer cases studied. Another study has suggested that chronic mycoplasma infection may contribute to prostate cancer development in BPH

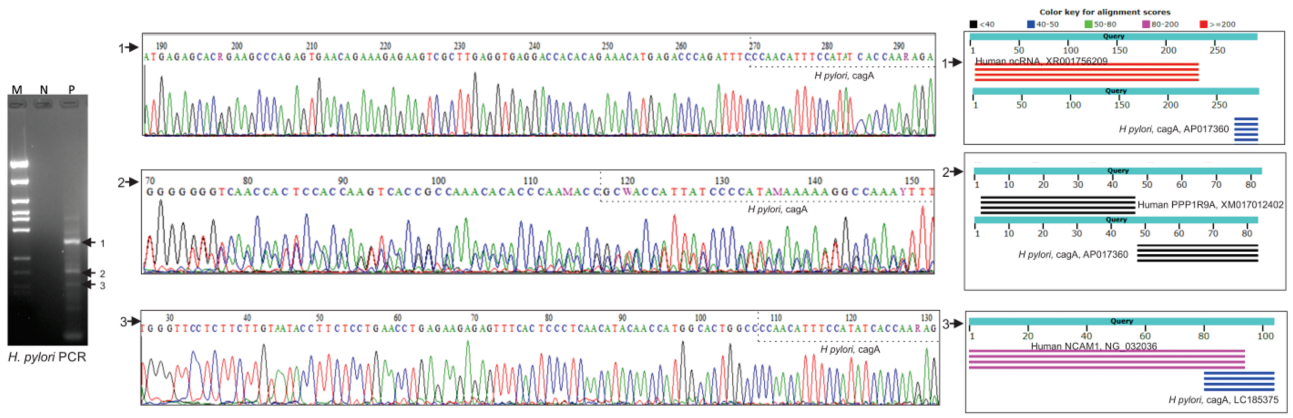

Figure 6. *Helicobacter cagA* genomic integrations detected during PCR validation of *Helicobacter* signature in the prostate cancer. The gel picture of EtBr-stained amplicons run on 2% agarose gel shows the three bands of interest (numbered). The electropherograms for the individual bands that were sequenced and the results of the NCBI BLAST program of the sequences are shown. The BLAST alignment shows that a part of the sequence of the three amplicons is bacterial and the rest is of human. On the gel picture, M is DNA ladder of RsaI digested φX/174, N is non-template control, P is pooled WTA product of prostate cancer samples.

cells (34). In this regard, we detected mycoplasma signatures with high hybridization signal intensity in at least 90% of the prostate cancer samples examined. Thus, there is good agreement between previous studies and our study as to the presence of bacterial signatures in prostate cancer.

We also detected signatures of *Helicobacter* in >90% of prostate cancer cases. In this regard, previous studies have suggested that *H.pylori* infection may contribute to prostate diseases (38,39). One of these studies demonstrated *H.pylori* DNA in the prostatic tissue of both BPH and a prostate cancer patient (39). Our study detected signals from few *Helicobacter* probes in the BPH controls, none of which were significantly higher than in the cancers (Supplementary Table S3, available at *Carcinogenesis* Online). Thus, *Helicobacter* is likely to be a low-level component of the BPH microbiome as suggested by the PCR validation (Figure 4). In the present study, we found sequences of *H.pylori* to be integrated at certain locations in the human somatic chromosomes 17, 7 and 11 (17q21.31, 7q21.3, 11q23.2). The integration of the *cagA* gene sequence in *PPP1R9A* and *NCAM1* gene locations may result in deregulation of their gene expression. Although *PPP1R9A* gene overexpression seen in prostate cancer (54) provides growth advantage to malignant cells, downregulation of *NCAM1* gene has been identified in several human cancers suggesting that it might function as a tumor repressor (55). It was thus interesting to find *H.pylori cagA* gene integrations in *PPP1R9A* and *NCAM1* genes, which may be a contributing factor to prostate tumorigenesis. Notably, one previous report has suggested *Helicobacter* DNA integration in a stomach adenocarcinoma (56). Although few studies have been done to examine the integration of bacterial sequences in human cell DNA, such integrations have been reported more frequently in tumors than in controls (56). Our study suggests that there is a marked increase in integration of viral and microbial sequences in prostate tumor DNA; we have reported similar findings for other tumors (14,15).

The integrated *H.pylori* DNA that we detected in prostate tumor cells include the sequences of the *cagA* gene, which encodes the immune-dominant *cagA* virulence factor (57). CagA is also associated with more severe gastric cancer (57–61). Gastric cancer patients are at least twice as likely to be infected with an *H.pylori* strain that is *cagA* positive than one that is *cagA* negative (59,61). This is significant because cagA is known to activate proto-oncogenes and inactivate tumor suppressor genes (33,62); thus, cagA plays an important role in disease progression in cases of gastric cancer (57,63). Thus, the finding of *cagA* sequences integrated in prostate cancer cell DNA poses the intriguing possibility that it may function in the establishment or progression of the cancer.

Among fungi, dermatophytes comprised the largest number of the fungal signatures detected in the prostate cancer samples. This may be because they are commonly detected in cancer patients (64). Similarly, the abundant detection of yeasts in the cancer cases is consistent with studies showing that opportunistic yeast infections are common in cancer cases (65–67). Also, consistent with the present study are previous reports of high incidence of microsporidia, such as Encephalitozoon and Fonsecaea, in cancer (68,69). In particular, chronic chromoblastomycosis, caused by Fonsecaea, has been suggested to promote squamous cell carcinoma (69). As was the case with viral and bacterial signatures, there is substantial agreement between previous studies and our study as to the presence of fungal signatures in prostate cancer.

A surprising result from our study is the presence of parasite signatures in prostate tumor samples. However, parasites have been directly or indirectly associated with several different cancers (70–77). For example, *Anisakis* has been suggested to be a risk factor in colon and stomach cancers (76); *Toxoplasma* has been shown to induce prostatic inflammation and hyperplasia (75); *Blastocystis* is found predominantly in colorectal cancer (77); *Schistosoma* in bladder cancer (72); and *Strongyloides* has been associated with gastric and other cancers (71,74). In addition, the intestinal nematode *Anisakis* has been reported previously in the male urinary tract (78). We report finding signatures for these parasites in the prostate tumor samples. We also detected *Plasmodium* in the prostate tumor samples, which is interesting because it has been reported to activate EBV from latency (73). In addition, *Plasmodium* has been reported to be a potent mutagen that can indirectly induce chromosomal damage (79), produce reactive oxygen species (80) and inhibit apoptosis (81), potentially facilitating oncogenesis.

Hierarchical cluster analysis showed that the microbiome signatures of the prostate tumors could be grouped into distinct clusters (1, 2a and 2b), suggesting that within prostate tumors different microbiomes are present. Thus, the microbiome may correlate with diagnostic aspects of the disease. Using the limited clinical data that were available for de-identified samples (Gleason grades, Gleason scores and the reported stages of the cancer), we looked for correlative trends between the clinical data and the specific clusters. The sample size is small, and our findings are largely correlative; however, we did find correlations that suggest that specific microbiome signatures may have prognostic and/or diagnostic value. In this regard, we examined the correlations between specific viral and microbial signatures and Gleason score and stages of cancer. We found that certain signatures were significantly higher in prostate cancer with lower Gleason scores (Supplementary Table S5, available at *Carcinogenesis* Online), where other signatures were higher in prostate cancer with higher Gleason score (see Results; Figure 3C; Supplementary Table S5, available at *Carcinogenesis* Online). These finding suggest that hybridization intensity of a group of specific viruses and microorganisms can provide significant prognostic and diagnostic value. It is likely that a study of a larger number of samples will clarify and expand the number of distinct clusters and more closely align clinical data to specific clusters and specific signatures.

As in our previous PathoChip studies of tumor microbiota (14,15), we validated the specific Papillomaviridae and Herpesviridae signatures by sequencing the prostate tumor samples captured by hybridization to selected positive probes of PathoChip screen. Since the tumors are heterogeneous, we pooled samples for the probe-capture sequencing as we only wanted to validate the presence of those signatures in prostate cancer samples. The results validated the PathoChip results. Probably, the most intriguing result from the verification studies was the finding that some captured viral and microbial sequences contained flanking sequences that aligned to human sequences, thus suggesting sites of viral and microbial DNA integration in human chromosomes. We found many examples of viral and microbial integration in the tumor DNA suggesting that tumor cells exhibit greatly increased recombinatorial activity during the development and expansion of the tumor. We show specific hotspots for integration, which may perturb gene expression or miRNA/lncRNA function in ways that potentially modulate or potentiate oncogenesis.

The controls for the study were derived from patients with BPH since normal prostate samples are very rare. BPH is an

inflammatory pathologic condition of the prostate which in some cases could be caused by microbial infections, and may be a precursor to prostate cancer development (3,82). Thus, it is quite possible that the microbiome between BPH and cancerous tissue may be shared and that viral and bacterial integrations may occur during a pre-cancerous BPH condition. Using the same primers used to validate several microbial insertions in the prostate cancer (Figure 5, Supplementary Figure S7, available at *Carcinogenesis* Online), we also analyzed integration in BPH. These analyses showed similar amplicon from the prostate controls (Supplementary Figure S9, available at *Carcinogenesis* Online), which, when sequenced, confirmed these integrations in BPH and the cancer. Overall, the BPH and prostate tumor microbiomes may overlap; however, our data show that there is clearly more diverse microbiome in tumor.

Observing similar HPV18 and KSHV insertions in the BPH samples as in the cancers were not surprising, as inflammatory prostate of BPH patients were not devoid of those viral detections, although significantly lower than in the cancers (Supplementary Table S3, available at *Carcinogenesis* Online). However, we did perform quantitative RT-PCR on the affected genes to see whether the gene expression were different in the cancers compared with the controls (Supplementary Figure S10, available at *Carcinogenesis* Online). The host genes, in which microbial insertions were detected, are already known to be associated with oncogenesis (83–86), and the differential expression of those genes that we detected in the prostate cancer samples and in the controls (Supplementary Figure S10, available at *Carcinogenesis* Online) were also previously reported (83–86). This may or may not be directly related to microbial genomic insertions within those genes.

In conclusion, we have identified diverse microbiome signatures associated with prostate cancer samples. Many of the viruses and microorganisms we detected have previously been associated with prostate cancer or other cancers. Our observation of integrations of viruses and bacteria into both BPH and prostate cancer cells is the first demonstration of the diversity of viruses and microorganism that can integrate. The prevalence of integrations, especially in the cancer cells, suggests that these cells may have heightened recombinatorial activity. In several cases, the integrations of viral (HPV18, KSHV) and bacterial (*Helicobacter*) sequences potentially result in gene expression perturbations, which could influence the initiation or progression of the cancer. Finally, the hierarchical clustering analysis of the prostate tumor microbiome suggests that microbiome signatures may correlate with clinical data, suggesting that the signatures may provide biomarkers for diagnostic and prognostic purposes.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Funding

Avon Foundation for Women (Avon-02-2012-053 to E.S.R.); the Abramson Cancer Center Director's fund.

Acknowledgements

We acknowledge Drs Fang Chen and Dr Don Baldwin for technical assistance.
Conflict of Interest Statement: None declared.

References

1. Siegel, R.L *et al.* (2019). Cancer statistics, 2019. CA Cancer J Clin., 69(1), 7–34.
2. Nelson, W.G. *et al.* (2003) Prostate cancer. N. Engl. J. Med., 349, 366–381.
3. Sfanos, K.S. *et al.* (2012) Prostate cancer and inflammation: the evidence. Histopathology, 60, 199–215.
4. Fassi Fehri, L. *et al.* (2011) Prevalence of Propionibacterium acnes in diseased prostates and its inflammatory and transforming activity on prostate epithelial cells. Int. J. Med. Microbiol., 301, 69–78.
5. Shannon, B.A. *et al.* (2006) Links between *Propionibacterium acnes* and prostate cancer. Future Oncol., 2, 225–232.
6. Klein, E.A. *et al.* (2008) Inflammation, infection, and prostate cancer. Curr. Opin. Urol., 18, 315–319.
7. Das, D. *et al.* (2008) BK virus as a cofactor in the etiology of prostate cancer in its early stages. J. Virol., 82, 2705–2714.
8. Ahsan, N. *et al.* (2006) Polyomaviruses and human diseases. Adv. Exp. Med. Biol., 577, 1–18.
9. Samanta, M. *et al.* (2003) High prevalence of human cytomegalovirus in prostatic intraepithelial neoplasia and prostatic carcinoma. J. Urol., 170, 998–1002.
10. Zambrano, A. *et al.* (2002) Detection of human polyomaviruses and papillomaviruses in prostatic tissue reveals the prostate as a habitat for multiple viral infections. Prostate, 53, 263–276.
11. Whitaker, N.J. *et al.* (2013) Human papillomavirus and Epstein Barr virus in prostate cancer: koilocytes indicate potential oncogenic influences of human papillomavirus in prostate cancer. Prostate, 73, 236–241.
12. Banerjee, S. *et al.* (2015) Distinct microbiological signatures associated with triple negative breast cancer. Sci. Rep., 5, 15162.
13. Banerjee, S. *et al.* (2018) Distinct microbial signatures associated with different breast cancer types. Front. Microbiol., 9, 951.
14. Banerjee, S. *et al.* (2017) Microbial signatures associated with oropharyngeal and oral squamous cell carcinomas. Sci. Rep., 7, 4036.
15. Banerjee, S. *et al.* (2017) The ovarian cancer oncobiome. Oncotarget, 8, 36225–36245.
16. Yu, H. *et al.* (2015) Urinary microbiota in patients with prostate cancer and benign prostatic hyperplasia. Arch. Med. Sci., 11, 385–394.
17. Cavarretta, I. *et al.* (2017) The microbiome of the prostate tumor microenvironment. Eur. Urol., 20, 30250–30256.
18. Chen, Y. *et al.* (2015) Identification of pathogen signatures in prostate cancer using RNA-seq. PLoS One, 10, e0128955.
19. Baldwin, D.A. *et al.* (2014) Metagenomic assay for identification of microbial pathogens in tumor tissues. MBio, 5, e01714–e01714.
20. Corless, C.L. *et al.* (2012) Tackling formalin-fixed, paraffin-embedded tumor tissue with next-generation sequencing. Cancer Discov., 2, 23–24.
21. Schweiger, M.R. *et al.* (2009) Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS One, 4, e5548.
22. Ludyga, N. *et al.* (2012) Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. Virchows Arch., 460, 131–140.
23. R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
24. Kolde, R. (2015) Pheatmap: Pretty Heatmaps. R Package Version 1.0.2.
25. Charrad, M. *et al.* (2014) NbClust: an R package for determining the relevant number of clusters in a data set. J. Stat. Softw., 61, 1–36.
26. Wu, T.D. *et al.* (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics, 26, 873–881.
27. Liao, Y. *et al.* (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 30, 923–930.
28. Thorvaldsdóttir, H. *et al.* (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform., 14, 178–192.
29. Ho, D.W. *et al.* (2015) Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. Oncotarget, 6, 20959–20963.
30. Krämer, A. *et al.* (2014) Causal analysis approaches in ingenuity pathway analysis. Bioinformatics, 30, 523–530.

31. TA, H. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser., 41, 95–98.
32. Sayers, E.W. *et al.* (2017) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res., 45, D12–D17.
33. André, A.R. *et al.* (2010) Gastric adenocarcinoma and *Helicobacter pylori*: correlation with p53 mutation and p27 immunoexpression. Cancer Epidemiol., 34, 618–625.
34. Namiki, K. *et al.* (2009) Persistent exposure to mycoplasma induces malignant transformation of human prostate cells. PLoS One, 4, e6872.
35. Baralle, D. *et al.* (2005) Splicing in action: assessing disease causing sequence changes. J. Med. Genet., 42, 737–748.
36. Horikawa, I. *et al.* (2001) *cis*-Activation of the human telomerase gene (hTERT) by the hepatitis B virus genome. J. Natl Cancer Inst., 93, 1171–1173.
37. Li, X. *et al.* (2014) The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. J. Hepatol., 60, 975–984.
38. Al-Marhoon, M.S. (2008) Is there a role for *Helicobacter pylori* infection in urological diseases? Urol. J., 5, 139–143.
39. Al-Marhoon, M.S. *et al.* (2015) Molecular evidence of *Helicobacter pylori* infection in prostate tumors. Curr. Urol., 8, 138–143.
40. Lin, Y. *et al.* (2011) Human papillomavirus 16 or 18 infection and prostate cancer risk: a meta-analysis. Ir. J. Med. Sci., 180, 497–503.
41. Silverman, R.H. *et al.* (2010) The human retrovirus XMRV in prostate cancer and chronic fatigue syndrome. Nat. Rev. Urol., 7, 392–402.
42. Thompson, M.P. *et al.* (2004) Epstein-Barr virus and cancer. Clin. Cancer Res., 10, 803–821.
43. Fan, H. (2007) A new human retrovirus associated with prostate cancer. Proc. Natl Acad. Sci. USA, 104, 1449–1450.
44. Sarid, R. *et al.* (2011) Viruses and human cancer: from detection to causality. Cancer Lett., 305, 218–227.
45. Johal, H. *et al.* (2010) DNA of mouse mammary tumor virus-like virus is present in human tumors influenced by hormones. J. Med. Virol., 82, 1044–1050.
46. Smelov, V. *et al.* (2016) Detection of DNA viruses in prostate cancer. Sci. Rep., 6, 25235.
47. Taghavi, A. *et al.* (2015) Polyomavirus hominis 1 (BK virus) infection in prostatic tissues: cancer versus hyperplasia. Urol. J., 12, 2240–2244.
48. Schlaberg, R. *et al.* (2009) XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. Proc. Natl Acad. Sci. USA, 106, 16351–16356.
49. Yang, Y. *et al.* (2014) Microbial imbalance and intestinal pathologies: connections and contributions. Dis. Model. Mech., 7, 1131–1142.
50. Shen, X.J. *et al.* (2010) Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. Gut Microbes, 1, 138–147.
51. Shin, N.R. *et al.* (2015) Proteobacteria: microbial signature of dysbiosis in gut microbiota. Trends Biotechnol., 33, 496–503.
52. Kim, J. *et al.* (2002) Activation of toll-like receptor 2 in acne triggers inflammatory cytokine responses. J. Immunol., 169, 1535–1541.
53. Bielecki, R. *et al.* (2005) Subclinical prostatic inflammation attributable to *Chlamydia trachomatis* in a patient with prostate cancer. Med. Wieku Rozwoj., 9, 87–91.
54. Ribarska, T. *et al.* (2014) Deregulation of an imprinted gene network in prostate cancer. Epigenetics, 9, 704–717.
55. Suzuki, M. *et al.* (2017) Loss of expression of the neural cell adhesion molecule 1 (NCAM1) in atypical teratoid/rhabdoid tumors: a new diagnostic marker? Appl. Cancer Res., 37, 14.
56. Riley, D.R. *et al.* (2013) Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. PLoS Comput. Biol., 9, e1003107.
57. Jones, K.R. *et al.* (2010) A tale of two toxins: *Helicobacter pylori* CagA and VacA modulate host pathways that impact disease. Front. Microbiol., 1, 115.
58. Crabtree, J.E. *et al.* (1991) Mucosal IgA recognition of *Helicobacter pylori* 120 kDa protein, peptic ulceration, and gastric pathology. Lancet, 338, 332–335.
59. Gwack, J. *et al.* (2006) CagA-producing *Helicobacter pylori* and increased risk of gastric cancer: a nested case-control study in Korea. Br. J. Cancer, 95, 639–641.
60. Covacci, A. *et al.* (1993) Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. Proc. Natl Acad. Sci. USA, 90, 5791–5795.
61. Blaser, M.J. *et al.* (1995) Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res., 55, 2111–2115.
62. Meyer-ter-Vehn, T. *et al.* (2000) *Helicobacter pylori* activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes c-*fos* and c-*jun*. J. Biol. Chem., 275, 16064–16072.
63. Rieder, G. *et al.* (2005) *Helicobacter pylori* cag-type IV secretion system facilitates corpus colonization to induce precancerous conditions in Mongolian gerbils. Gastroenterology, 128, 1229–1242.
64. Irimie, M. *et al.* (2015) Prevalence and antifungal susceptibility patterns of dermatophytes isolated from patients with neoplastic diseases: a case control study. Acta Dermatovenerol. Croat., 23, 108–113.
65. Miceli, M.H. *et al.* (2011) Emerging opportunistic yeast infections. Lancet Infect. Dis., 11, 142–151.
66. Singh, G.K. *et al.* (2017) Spectrum of fungal infection in head and neck cancer patients on chemoradiotherapy. J. Egypt. Natl Canc. Inst., 29, 33–37.
67. Krcmery, V. *et al.* (1999) Invasive yeast infections other than *Candida* spp. in acute leukaemia. J. Hosp. Infect., 41, 181–194.
68. Lono, A.R. *et al.* (2008) Incidence of microsporidia in cancer patients. J. Gastrointest. Cancer, 39, 124–129.
69. Azevedo, C.M. *et al.* (2015) Squamous cell carcinoma derived from chronic chromoblastomycosis in Brazil. Clin. Infect. Dis., 60, 1500–1504.
70. Krueger, H. *et al.* (2010) HPV and Other Infectious Agents in Cancer: Opportunities for Prevention and Public Health. Oxford University Press. New York.
71. Seo, A.N. *et al.* (2015) Comorbid gastric adenocarcinoma and gastric and duodenal *Strongyloides stercoralis* infection: a case report. Korean J. Parasitol., 53, 95–99.
72. Srougi, V. *et al.* (2017) Carcinosarcoma of the bladder following local schistosomiasis infection. BMJ Case Rep., 2017, bcr2016218642.
73. van Tong, H. *et al.* (2017) Parasite infection, carcinogenesis and human malignancy. EBioMedicine, 15, 12–23.
74. Zueter, A.M. *et al.* (2014) Detection of *Strongyloides stercoralis* infection among cancer patients in a major hospital in Kelantan, Malaysia. Singapore Med. J., 55, 367–371.
75. Colinot, D.L. *et al.* (2017) The common parasite *Toxoplasma gondii* induces prostatic inflammation and microglandular hyperplasia in a mouse model. Prostate, 12, 23362.
76. Garcia-Perez, J.C. *et al.* (2015) Previous exposure to the fish parasite *Anisakis* as a potential risk factor for gastric or colon adenocarcinoma. Medicine, 94, e1699.
77. Mohamed, A.M. *et al.* (2017) Predominance and association risk of *Blastocystis hominis* subtype I in colorectal cancer: a case control study. Infect. Agent. Cancer, 12, 21.
78. Zahariou, A. *et al.* (2007) Enterobius vermicularis in the male urinary tract: a case report. J. Med. Case Rep., 1, 137.
79. Robbiani, D.F. *et al.* (2015) Plasmodium infection promotes genomic instability and AID-dependent B cell lymphoma. Cell, 162, 727–737.
80. Eze, M.O. *et al.* (1990) Reactive oxygen production against malaria – a potential cancer risk factor. Med. Hypotheses, 32, 121–123.
81. Carmen, J.C. *et al.* (2007) Suicide prevention: disruption of apoptotic pathways by protozoan parasites. Mol. Microbiol., 64, 904–916.
82. Sfanos, K.S. *et al.* (2013) Infections and inflammation in prostate cancer. Am. J. Clin. Exp. Urol., 1, 3–11.
83. Yeh, H.Y. *et al.* (2009) Identifying significant genetic regulatory networks in the prostate cancer from microarray data based on transcription factor analysis and conditional independency. BMC Med. Genomics, 2, 70.
84. Vázquez-Arreguín, K. *et al.* (2016) The Oct1 transcription factor and epithelial malignancies: old protein learns new tricks. Biochim. Biophys. Acta, 1859, 792–804.
85. Marotti, J.D. *et al.* (2017) P-Rex1 expression in invasive breast cancer in relation to receptor status and distant metastatic site. Int. J. Breast Cancer, 2017, 4537532.
86. Kanwal, M. *et al.* (2018) MUC16 overexpression induced by gene mutations promotes lung cancer cell growth and invasion. Oncotarget, 9, 12226–12239.