Identification of Pathogen Signatures in Prostate Cancer Using RNA-seq

Yunqin Chen, Jia Wei*

AstraZeneca, R&D Information, 199 Liangjing Road, Zhangjiang Hi-Tech Park, Shanghai, 201203, China

* Jenny.wei@astrazeneca.com

Abstract

Infections of the prostate by bacteria, human papillomaviruses, polyomaviruses, xenotropic murine leukemia virus (MLV)-related gammaretroviruses, human cytomegaloviruses and other members of the herpesvirus family have been widely researched. However, many studies have yielded conflicting and controversial results. In this study, we systematically investigated the transcriptomes of human prostate samples for the unique genomic signatures of these pathogens using RNA-seq data from both western and Chinese patients. Human and nonhuman RNA-seq reads were mapped onto human and pathogen reference genomes respectively using alignment tools Bowtie and BLAT. Pathogen infections and integrations were analyzed in adherence with the standards from published studies. Among the nine pathogens (Propionibacterium acnes, HPV, HCMV, XMRV, BKV, JCV, SV40, EBV, and HBV) we analyzed, Propionibacterium acnes genes were detected in all prostate tumor samples and all adjacent samples, but not in prostate samples from healthy individuals. SV40, HCMV, EBV and low-risk HPVs transcripts were detected in one tumor sample and two adjacent samples from Chinese prostate cancer patients, but not in any samples of western prostate cancer patients; XMRV, BKV and JCV sequences were not identified in our work; HBV, as a negative control, was absent from any samples. Moreover, no pathogen integration was identified in our study. While further validation is required, our analysis provides evidence of Propionibacterium acnes infections in human prostate tumors. Noted differences in viral infections across ethnicity remain to be confirmed with other large prostate cancer data sets. The effects of bacterial and viral infections and their contributions to prostate cancer pathogenesis will require continuous research on associated pathogens.

Introduction

As the second most common cause of cancer-related death among men [1], prostate cancer (PCa) remains a great public health concern. A significant amount of evidence has revealed that chronic inflammation can be associated with the onset of PCa [2, 3]. Chronic inflammatory infiltrates are common findings in prostate tissue samples and pathogen infections are considered to be one possible cause thereof.

It is known that Propionibacterium acnes (*P. acnes*) play an important role in human health and disease [4]. *P. acnes* in skin generally has a positive effect on human health by preventing the colonization of pathogenic microorganisms, but when the host becomes compromised (trauma, injury or alterations in immune status), it can display pathogenic potential [5, 6]. The presence of *P. acnes* has been strongly correlated with histological inflammation, suggesting that this bacterium might be linked to cancer development [7, 8]. Several studies have demonstrated the high prevalence of the Gram-positive bacterium *P.acnes* in the prostate tissues of men diagnosed with prostate disease [9, 10]. Infection of a prostate epithelial cell line with *P. acnes* induces a strong host inflammatory response and transformation, which could be a trigger for cancer initiation or progression [6, 10].

Viruses cause 10–15% of all human cancers. The association between infections caused by DNA viruses and the development of tumors is well established in many cancer types. Viruses associated with human cancers are known as 'tumor viruses'. Most of these viruses are capable of integrating into the host genome and immortalizing the target cell in order to facilitate their own replication. The infected cell expresses the viral genes, which are able to induce cell growth, proliferation and prevent apoptosis. For example, a high hepatitis B virus (HBV) load and chronic hepatitis B (CHB) infection increases the risk of developing hepatocellular carcinoma. HBV is a DNA virus that can integrate DNA into the host genome thereby increasing the yield of transactivator protein HBxAg. HBxAg is involved in many metabolic pathways [11]. Recently, some researchers have identified recurrent HBV integration events at the known and putative cancer-related genes such as TERT, MLL4 and CCNE1. These genes demonstrated upregulated gene expression in tumors but not in normal tissues [12]. The elucidation of tumor-DNA virus associations will enhance our fundamental knowledge of oncogenesis mechanisms and provide a foundation for cancer prevention initiatives.

More and more research has indicated that viral infections may lead to chronic or recurrent inflammation of the prostate and even initiate or promote carcinogenesis [13–15]. Viral products are able to interact with the interferon signalling pathway and induce cell transformation synergistically [16]. A number of viruses are reported to be associated with prostate cancer or prostate infections, namely Human papillomaviruses (HPVs), polyomaviruses (BK, JC, and SV40), and members of the herpesvirus family (HCMV, EBV) [17–21].

HPV is now recognized as one of the major causes of cervical cancer [22]. High-risk HPVs have also been frequently identified in both benign and malignant prostate tissues [23]. There are more than 100 different types of HPV and over 30 of these types are transmitted sexually, making HPV the most common sexually transmitted disease. The various HPV types are divided into two super categories- those that are more likely to develop into cancer and those that are less likely. The so-called "high-risk" forms are more likely to lead to the development of cancer, while "low-risk" viruses rarely develop into cancer.

HCMV infection typically goes unnoticed in healthy people, though it can be life-threatening for the immuno-compromised, such as HIV-infected persons, organ transplant recipients, or infants. After infection, HCMV has the ability to remain latent within the body over long periods. Eventually, it may cause mucoepidermoid carcinoma and possibly other malignancies. It is also reported that HCMV may be associated with prostate cancer [19, 24]. The Epstein-Barr virus (EBV), best known as the cause of infectious mononucleosis, is implicated in some malignant lymphomas and lymphoepithelioma-like carcinomas [25].

Polyomaviruses are small (40–50 nanometers in diameter) non-enveloped DNA viruses, and icosahedral in shape. They are potentially oncogenic (tumor-causing); and often persist as latent infections in a host without causing diseases, but may cause tumors in a host of a different species, or a host with an ineffective immune system. The polyomaviruses that infect

humans, including the BK viruses (BKV), JC viruses (JCV), and simian virus 40 (SV40), typically cause infections that are subclinical and persistent [18].

The Xenotropic murine leukemia virus-related virus (XMRV) was first discovered in patients with PCa and later in patients with chronic fatigue syndrome (CFS) [26, 27]. Some further studies have provided evidence of XMRV infection in PCa [28–31] though its association with CFS and PCa has been largely discredited [32, 33]. Recent studies have suggested that the presence of XMRV may be a result of mouse DNA contamination [34, 35]. Because some results which indicate the presence of XMRV in PCa cannot be wholly attributed to sample contamination [36], here we examined the possible linkage between XMRV and PCa.

Uncovering the genomic effects of known pathogens in PCa remains a challenge. While most PCa research has focused on the PCR-based targeted detection of viruses, the advancement of next generation sequencing technology makes whole genome interrogation possible. Our goal is to analyze RNA-seq data from Chinese and western PCa patients in order to identify pathogens and their integration within the host genomes.

Materials and Methods

Three RNA-seq data sets

Data set 1: single-end miRNA-seq data of a pooled biopsy-proven PCa sample and a pooled control sample of Australian patients without detectable cancer. Small RNA sequencing was used to profile and compare miRNAs in the non-sperm cellular fraction of seminal fluid from men with biopsy-proven cancer (a pooled sample from 6 men) and men with elevated serum PSA but negative biopsy results (a pooled sample from 6 men) [37]. RNA was extracted using the Trizol reagent (Life Technologies) and cleaned up using RNeasy Mini kits (Qiagen). The single-end miRNA-seq data was downloaded from EBI ENA (http://www.ebi.ac.uk/ena/data/view/SRP041082).

Data set 2: paired-end mRNA-seq data of PCa tissues from Caucasian patients. Transcriptomes (polyA+) of 20 prostate tumors and 10 matched adjacent tissues were sequenced using the Illumina GAII platform. RNA was extracted from samples using the Ribopure kit (Ambion). Total RNA samples were processed for transcriptome sequencing using the Illumina mRNA-seq protocol. The pathological status of tumor samples was confirmed before processing, and the tumor samples had a tumor cell percentage >80% with Gleason scores ranging from 6 to 9 [38]. The mRNA-seq data was downloaded from EBI ENA (http://www.ebi.ac.uk/ena/data/view/SRX022060-SRX022089).

Data set 3: paired-end mRNA-seq data of Chinese PCa tissues. Prostate cancer and adjacent normal tissues from 14 patients obtained from the Shanghai Changhai Hospital were used as a cohort for RNA sequencing (using Illumina HiSeq 2000 sequencing machine). The tumor samples had Gleason scores ranging from 4 to 8 [39]. The mRNA-seq data was downloaded from EBI ENA (http://www.ebi.ac.uk/ena/data/view/ERP000550). Five samples (4T, 5T, 6N, 11T, 12N) had no available paired-end reads, which we analyzed using a single-end strategy.

All the data sets were generated in the FASTQ format. The sample collection, sequencing and clinical information of the 3 RNA-seq data sets was listed in S1 File. The average read length for Data set 1, 2, 3 is 49bp, 36bp, and 90bp respectively.

Pathogens investigated in our work

We did a systemic literature review of viruses and bacteria reported in prostate cancer and finally selected seven viruses: HPV, HCMV, XMRV, BKV, JCV, SV40, EBV and one bacterium-Propionibacterium acnes (*P. acnes*) to investigate their pathogen associations with PCa. We

also used the HBV virus, which can cause hepatitis B and rarely occurs in patients with PCa as a negative control virus to evaluate our sequencing analysis results.

Reference Sequences used for mapping

The reference sequences consisted of human and pathogen sequences. Human genome and transcriptome sequences were downloaded from the NCBI website (NCBI build 37). Propionibacterium acnes and all the viral genome and transcriptome sequences were also taken from the NCBI build 37.

Pathogen detection and analysis of pathogen integration sites

mRNA-seq data analysis pipeline for Data set 2 and Data set 3. We used the NGS QC Toolkit (v2.3) [40] to eliminate bad mRNA-seq reads. Two criteria were used to select good reads: cutOffQualScore = 20 and cutOffReadLen4HQ = 90, meaning that 90% of the bases of a qualified read must have quality scores $> = 20$. All subsequent analyses were based on clean mRNA-seq reads.

Both Bowtie [41] and the BLAST-like alignment tool-BLAT [42] were used in our virus detection pipeline. Bowtie is an ultrafast read aligner which can quickly map tens of millions of single-end or paired-end reads. We used Bowtie2 (version 2.1.0) and applied default parameters to carry out the alignment.

BLAT is an alignment tool like BLAST, but structured differently. It can efficiently map short reads across exon—exon junctions and identify novel splice junctions. Here we use stand-alone BLAT v.35. The parameters applied to align reads with reference sequences are as follows: minScore = 20, minIdentity = 88, stepSize = 5, and combined—fine and no—fine mode.

Before mapping the reads, we performed an alignment between human and pathogen sequences to identify consensus sequences, which were used to filter false fusions. As shown in Fig 1A, raw reads were first mapped to human and pathogen reference sequences with Bowtie2, and then unmapped reads (only for length (read) >30) were picked out to perform local alignment with BLAT. If two paired-end reads were uniquely mapped with one read mapped to human sequences and the other read mapped to specific pathogen sequences, the paired-end reads were considered a human-pathogen fusion candidate read. If a read from one paired-end reads was uniquely mapped to human or pathogen sequences, and the other end read was uniquely from two parts: one part mapped to human sequences and the other part to pathogen sequences, it was labeled a raw human-pathogen fusion event. After detecting all the raw fusion events, we applied our criteria to filter out false positive reads and select fusion candidates.

For every single-end read with one part (P1) mapped to human sequences and the other part (P2) mapped to pathogen sequences (see Fig 1B), the following criteria were used for fusion filtering:

1. It is required that P1 is strictly mapped to human sequences and P2 is strictly mapped to pathogen sequences. Neither P1 nor P2 may overlap with consensus sequences.

2. The mapped ratio P1/(P1+P2) or P2/(P1+P1) $< = 0.8$.

3. P1 or P2 should be uniquely mapped.

4. P1 or P2 should not have low-complexity sequences.

After getting the fusion candidate reads, we mapped them to the human-pathogen consensus sequences to remove false positive fusions in cases where the fusion reads likely came from consensus sequences. Finally, we also manually checked their reliability by browsing their
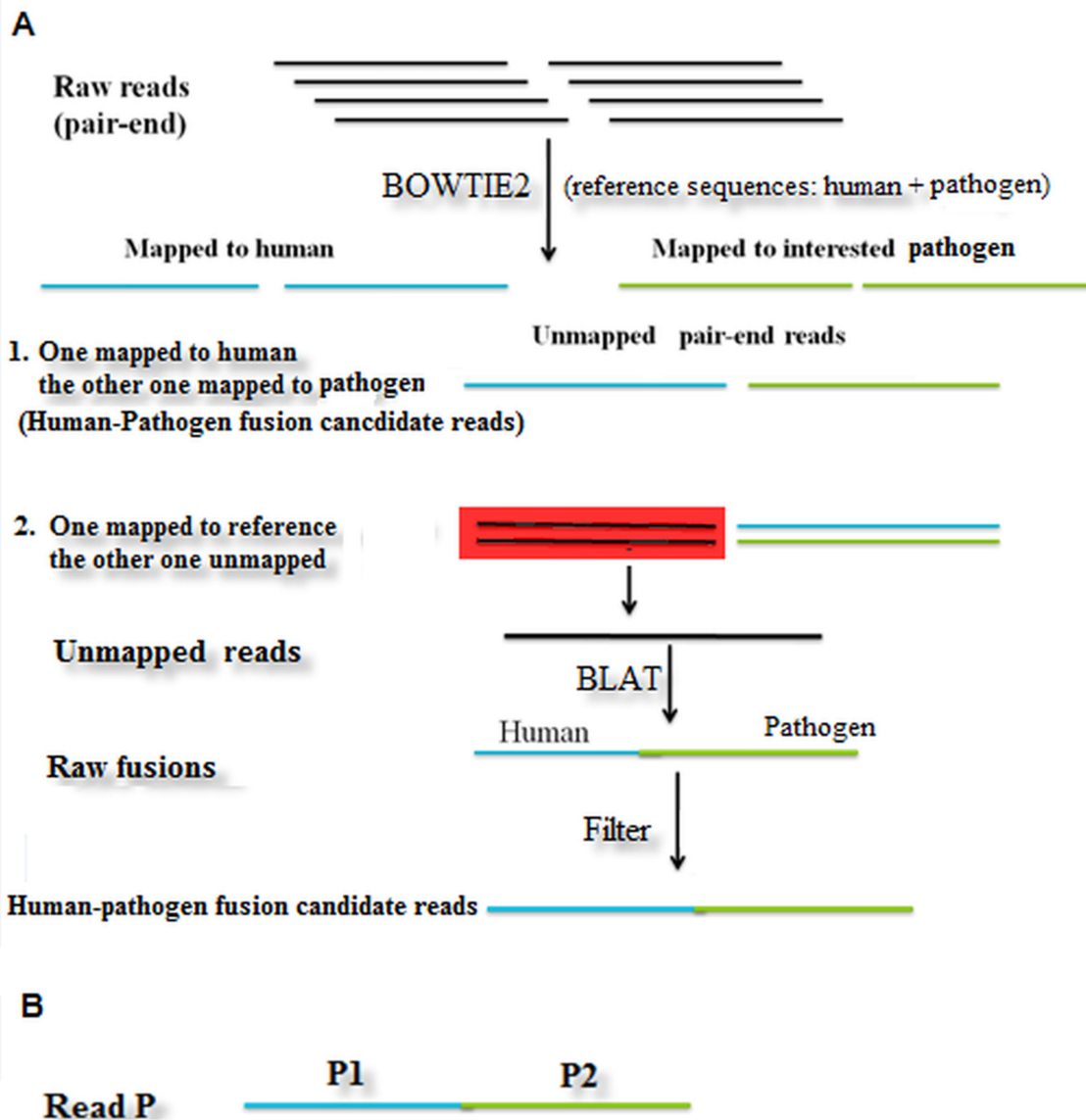
Fig 1. (A) Pathogen detection pipeline (B) Fusion Reads with one part (P1) mapped to human sequences and the other part (P2) mapped to pathogen sequences. Reads marked in blue are mapped to human sequences, reads marked in green are mapped to pathogen sequences, reads marked in black are unmapped.

doi:10.1371/journal.pone.0128955.g001

alignment using the web tool BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to ensure that the P1 and P2 regions were mapped onto human and pathogen sequences.

MiRNA-seq data analysis pipeline for Data set 1. MiRNA-seq is a type of RNA-seq, which uses next-generation sequencing technology to sequence MicroRNAs. First, we filtered out the raw reads with a base quality score of less than 20. Then we removed the 3' sequencing adapter using a BioPerl script (available at http://www.bioperl.org/wiki/Removing_sequencing_adapters). There were small RNAs like degraded mRNA fragments, rRNAs, tRNAs, miRNAs, siRNAs in the cleaned reads.

Because our purpose is to detect whether pathogens exist in the samples using the miRNA-seq data, we analyzed the clean reads using the mRNA-seq pipeline for single-end reads.

Measuring pathogen representation. In our study, quantified pathogen representation is determined by a measure of the overall count of mapped reads onto the pathogen genome and the expressed pathogen mRNAs in human samples.

We calculated the expressed pathogen mRNA levels by excluding non-transcribed pathogen genome elements. This eliminates and reduces the potential of nonsense reads and the number of non-transcribed pathogen genomic elements from the data pool. The transcript level and gene level counts were calculated and FPKM (fragments per kilobase of exon per million fragments mapped) were normalized using Cufflinks [43]. Then an FPKM filtering cutoff of 1.0 [44] or transcript count > = 2 was used to determine the level of expressed transcripts. Any pathogen expression level below the cutoff was labeled as having no obvious mRNA expression.

Results

We analyzed three RNA-seq data sets of PCa samples- two from the western population and one from the Chinese population- using the methods outlined in the Materials and Methods section of this research paper. For data-set 1 of the miRNA seq data, 1% of approximately 100M raw reads were mapped to the human genome (because most of the reads should be mapped to the human miRNA library). For data-set 2, consisting of western prostate patients, the effective alignment count was about 17M on average. About 86% of reads were mapped to the human genome. For data-set 3, consisting of Chinese prostate patients, the effective alignment count was approximately 56M on average. About 85% of reads were mapped to the human genome.

No virus but *P. acnes* detected in the PCa sample from Data set 1

None of the viruses were detected in either the pooled cancer sample or the normal sample from the western population, however *P. acnes* was identified in the cancer sample (see S2 File and Table 1). There were not high read loads of the *P. acnes* genome (about 77 reads mapped to each 100 kb on average) because of the limited number of small RNA sequences mapped to the reference sequences. The expressed genes with FPKM >1 were listed in Table 1. Intracellular protease / general stress protein 18 (YP_056816.1) responds to environmental stressors including extremes of temperature, exposure to toxins, and mechanical damage. Hypothetical protein PPA0381 (YP_055091.1) is likely to play a role in host-tissue biological degradation and inflammation [45]. Bacterial ABC transporters are essential in cell viability, virulence, and pathogenicity [46]. Fe(III) dicitrate ABC transporter (YP_056465.1) is participating in Iron ABC uptake systems which are important effectors of virulence.

Table 1. *P. acnes* genes expressed in Data set 1.

Gene id	Protein id	Protein name	FPKM for cancer sample (n = 6)	FPKM for control sample (n = 6)
gene137	YP_054855.1	reductase, ferredoxin	5.771	0
gene258	YP_054973.1	inorganic pyrophosphatase	12.5813	0
gene377	YP_055091.1	hypothetical protein PPA0381	18.245	0
gene558	YP_055264.1	hypothetical protein PPA2401	15.9918	0
gene1198	YP_055893.1	hypothetical protein PPA1186	8.3743	0
gene1397	YP_056085.1	imidazole glycerol phosphate synthase subunit HisF	10.5343	0
gene1795	YP_056465.1	Fe(III) dicitrate ABC transporter, ATP-binding protein	31.7288	0
gene2158	YP_056816.1	intracellular protease / general stress protein 18	23.8085	0
gene2186	YP_056844.1	hypothetical protein PPA2180	12.3472	0

doi:10.1371/journal.pone.0128955.t001

Table 2. Cancer-specific *P. acnes* genes in Data set 2: Western prostate cancer samples compared with matched adjacent samples.

Gene id	Protein id	Protein name	# Cancer sample
gene114	YP_054834.1	potassium-transporting ATPase subunit B	3
gene243	YP_054960.1	DNA topoisomerase I	4
gene460	YP_055173.1	dehydrogenase, myo-inositol 2-dehydrogenase	4
gene812	YP_055517.1	excinuclease ABC subunit B	3
gene1052	YP_055753.1	major facilitator superfamily permease	5
gene1592	YP_056275.1	hypothetical protein PPA1574	3
gene1887	YP_056556.1	elongation factor G	3

Only expressed transcripts with an FPKM > = 1.0 in at least 3 out 10 cancer samples are shown.

doi:10.1371/journal.pone.0128955.t002

Since the read length was < 60bp, we cannot use a single-end strategy to find human-bacteria fusions in these samples.

No virus but *P. acnes* detected in both PCa and adjacent benign samples in Data set 2

There were no obvious virus-mapped reads detected in the 20 PCa samples nor in the 10 matched adjacent samples from western patients. However the *P. acnes* signature was detected in both cancer samples and adjacent samples (see S2 File). 14 out of 20 cancer samples and 9 out 10 adjacent benign samples had at least 2 expressed *P. acnes* transcripts.

Using 10 paired cancer and adjacent samples, we identified 7 cancer-specific genes from *P. acnes*, all of which were expressed in at least 3 out 10 cancer samples (see Table 2). Among them, bacterial topoisomerase I (YP_054960.1) is required for preventing the hypernegative supercoiling of DNA during transcription and plays an important role in the transcription of stress genes during the bacterial stress response [47]. And elongation factor G (YP_056556.1) promotes the translocation step in bacterial protein synthesis.

We tried to find human-bacteria fusion in *P. acnes* identified samples, but no such paired-end reads where one read mapped to human sequences and the other mapped to *P. acnes* sequences were identified.

Viruses and *P. acnes* identified in cancer and adjacent samples in Data set 3

In this study, there were 9 matched cancer and adjacent samples, 3 unpaired cancer samples and 2 unpaired adjacent samples available in Data set 3 (see Materials and Methods).

The reads from *P. acnes* were identified in all Chinese cancer samples and adjacent normal samples. Sample 7N, 7T, 8N had a high RNA level of *P. acnes* (see S2 File). All but 2 cancer samples have evidence of expressed *P. acnes* transcripts. However, we did not find cancer-specific *P. acnes* transcripts which were expressed in at least 3 cancer samples as our standard threshold.

Interestingly, some virus reads were detected in 3 Chinese prostate samples-cancer sample 7T and adjacent samples 7N, 8N- all from patients with stage III PCa and with high PSA scores (7T&7N: PSA 30.33; 8N: PSA10.4) (see S1 File).

All 3 samples were identified as having viral DNA loads, expressed transcripts and more than one kind of virus was detected in each. The infected virus profiles were similar across the 3 samples (see Table 3). The viruses with more reads were SV40, HCMV, EBV and low-risk HPVs (like HPV49, 50, 88,108) (see Table 3 and S3 File).

Table 3. Viral transcripts detected in Chinese prostate samples.

Virus genome	Virus name	Expressed gene #	7N. read counts	7T. read counts	8N. read counts
NC_009334.1	EBV type 2	39	367	902	592
NC_006273.2	HCMV	38	2079	1916	2795
NC_007605.1	EBV	30	3224	2735	3637
NC_001669.1	SV40	10	1450	916	1470
NC_001591.1	HPV49	4	39	65	70
NC_001595.1	HPV7	4	15	51	17
NC_008188.1	HPV103	4	26	11	45
NC_010329.1	HPV88	4	177	79	151
NC_001691.1	HPV50	3	176	127	171
NC_004104.1	HPV90	3	3	11	6
NC_004500.1	HPV92	2	21242	21970	14097
NC_005134.2	HPV96	3	17	19	34
NC_008189.1	HPV101	3	6	6	11

Expressed gene # column shows the number of expressed genes in each virus detected sample from RNA-seq data.

7N. read counts column shows the number of paired-end reads mapped to each viral reference sequence for sample 7N.

7T. read counts column shows the number of paired-end reads mapped to each viral reference sequence for sample 7T.

8N. read counts column shows the number of paired-end reads mapped to each viral reference sequence for sample 8N.

doi:10.1371/journal.pone.0128955.t003

Similarly, we used the fusion detection pipeline to identify human-pathogen fusion events. We filtered the raw fusions using consensus sequences between each pathogen and human sequence. No viral or bacterial fusions were identified in Data set 3.

It is known that different *P. acnes* strains are associated with the prostate than those commonly found on the skin [48]. The draft genome sequences of two strains of *P. acnes* isolated from radical prostatectomy specimens (*P. acnes* P6 and *P. acnes* PA2) were reported [49]. We tried to use the genome sequences of 3 strains: KPA171202 (from skin), P6 and PA2 (from prostate), but the read counts were not high enough to detect the strain specific sequences of genes such as the housekeeping gene (RecA) or the putative hemolysin gene (Tly) in order to investigate the phylogenetic relationship between bacterial organisms [50, 51]. 16S rRNA is the gold standard for phylogenetic analysis. However, 16S rRNA could be removed during sequencing library preparation according to the protocols of RNA-seq technology. We did not identify any reads mapped to 16S rRNA.

Discussion

Microorganism infections of the prostate by *P. acnes*, HPV, HCMV, JCV and BKV and. the recently discovered, XMRV have been suggested as important risk factors in the development of PCa [17, 19, 23, 29, 52, 53]. Until now, studies aimed at replicating the associations between these pathogen infections with PCa have yielded conflicting results across different ethnic groups. Our current study analyzed associations between pathogen infections among Chinese and western PCa patients using RNA-seq. To our knowledge, this is the first analysis of the role of these viruses and bacteria in the development of prostate tumors of both western and Chinese patients using NGS data.

The bacterium *P. acnes* was detected in all three data sets, in both cancer tissues and adjacent benign prostate tissues except in normal samples from healthy individuals. Due to data limitations, we cannot rule out the possibility that the *P. acnes* we discovered in the prostate tissue may have come from the skin. Our observation was consistent with previous reports that

P. acnes were prevalent in both benign and malignant prostate tissues [9, 54]. It is highly likely that *P. acnes* were present in prostate cancer but not normal samples. *P.acnes* is believed to induce and maintain an inflammatory response by producing chemotactic factors that attract neutrophils [55] and the ability of *P. acnes* to survive phagocytosis in vitro has been confirmed [56]. Moreover, *P. acnes* releases proteases and other enzymes that contribute to tissue injury [45]. The host response to *P. acnes* is characterized by the production of proinflammatory cytokines such as tumour necrosis factor-α (TNF-α), interleukin 1-α (IL-1α) and interleukin 8 (IL-8) [57]. *P.acnes* has been shown to induce these cytokines in both macrophages and keratinocytes through the TLR2 and TLR4 signaling pathways [58, 59]. *P. acnes* may pose an increased risk for PCa given its role in inflammatory processes.

Another interesting finding in our study is that we observed the difference between adjacent tissues surrounding tumors and normal tissues from healthy individuals. The *P. acnes* did not exist in normal samples from healthy individuals but appeared in all adjacent samples of PCa patients in our work. The tissues adjacent to prostate tumors may also be undergoing tumor related changes, therefore careful characterization of these different tissues is necessary to understand the molecular changes leading up to PCa [60].

No viruses were detected in western prostate samples in Data set 1 and 2. But viruses like HCMV, EBV, SV40 and low-risk HPVs were identified in some of the Chinese prostate tissues (both cancer and adjacent samples) in Data set 3. The presence of both HPV and EBV gene sequences were present in an equal proportion of normal, benign, and PCa specimens from Australian males. However, using a range of analytical techniques including the in situ polymerase chain reaction (IS-PCR) and the standard liquid PCR before finally sequencing the product [21], no DNA virus transcripts were detected in the PCa samples using RNA-seq data from the TCGA Portal (https://tcga-data.nci.nih.gov/tcga/) [61, 62]. The different results from different studies are most likely attributable to sample differences: ethnic difference, pathological difference, etc. The Chinese sample 7N, 7T, 8N were infected by both *P. acnes* and other viruses. The presence of multiple viral sequences in the same prostate tumor sample is particularly noteworthy because these observations confirm the assessment of Zambrano et al.[63] that the prostate is a habitat for multiple viral and other infections, some of which may have oncogenic potential. Because only 2 out of 14 patients (patient No.7 and patient No.8) in Data set 3 had viral infections, no solid conclusion can be drawn as to whether those viruses contribute to the development of PCa.

XMRV and other viruses were not detected in any of the three datasets. Our results are consistent with the majority of published studies on XMRV, which show that XMRV is not present in humans [33, 64–66]. The absence of high risk HPV16 and HPV18 from any of the data sets analyzed indicates that high risk HPV is not associated with PCa in the investigated patients.

It has been reported that viral and bacterial integration occurs in the human somatic genome and may play a role in carcinogenesis. The fact that we did not detect any human-bacteria or human-virus fusion in any data set may suggest that *P. acnes* and the 4 viruses (HCMV, EBV, SV40, low- risk HPVs) may not pose severe risks for the development of prostate cancer.

Our study yielded a number of interesting findings pertaining to the pathogen signatures of PCa patients. The major limitations of this study are the small RNA-seq data set size and the lack of available tissue samples for validation. While the findings of this study, especially the presence of viruses in Chinese PCa patients, remain to be validated by large prostate data sets, our analysis sheds light on the emerging application of RNA-seq technology in detecting pathogen signatures in human cancers and elucidating novel mechanisms of pathogen-driven cancer pathogenesis.

Supporting Information

S1 File. Prostate sample information for the three data sets.
(XLSX)

S2 File. Mapped read coverage of *P. acnes* sequences in the three data sets.
(XLSX)

S3 File. Viral genes expressed in samples 7N, 7T, 8N of Data set 3.
(XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: YQC JW. Performed the experiments: YQC JW. Analyzed the data: YQC. Contributed reagents/materials/analysis tools: YQC. Wrote the paper: JW YQC.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: a cancer journal for clinicians. 2013; 63(1):11–30. doi: 10.3322/caac.21166 PMID: 23335087.

2. Nelson WG, De Marzo AM, Isaacs WB. Prostate cancer. The New England journal of medicine. 2003; 349(4):366–81. doi: 10.1056/NEJMra021562 PMID: 12878745.

3. Sfanos KS, De Marzo AM. Prostate cancer and inflammation: the evidence. Histopathology. 2012; 60(1):199–215. doi: 10.1111/j.1365-2559.2011.04033.x PMID: 22212087

4. McDowell A, Patrick S, Eishi Y, Lambert P, Eady A. Propionibacterium acnes in human health and disease. BioMed research international. 2013; 2013:493564. doi: 10.1155/2013/493564 PMID: 24455697; PubMed Central PMCID: PMC3884696.

5. Bojar RA, Holland KT. Acne and Propionibacterium acnes. Clinics in dermatology. 2004; 22(5):375–9. doi: 10.1016/j.clindermatol.2004.03.005 PMID: 15556721.

6. Perry A, Lambert P. Propionibacterium acnes: infection beyond the skin. Expert review of anti-infective therapy. 2011; 9(12):1149–56. doi: 10.1586/eri.11.137 PMID: 22114965.

7. Palapattu GS, Sutcliffe S, Bastian PJ, Platz EA, De Marzo AM, Isaacs WB, et al. Prostate carcinogenesis and inflammation: emerging insights. Carcinogenesis. 2005; 26(7):1170–81. doi: 10.1093/carcin/bgh317 PMID: 15498784.

8. Cohen RJ, Shannon BA, McNeal JE, Shannon T, Garrett KL. Propionibacterium acnes associated with inflammation in radical prostatectomy specimens: a possible link to cancer evolution? The Journal of urology. 2005; 173(6):1969–74. doi: 10.1097/01.ju.0000158161.15277.78 PMID: 15879794.

9. Fassi Fehri L, Mak TN, Laube B, Brinkmann V, Ogilvie LA, Mollenkopf H, et al. Prevalence of Propionibacterium acnes in diseased prostates and its inflammatory and transforming activity on prostate epithelial cells. International journal of medical microbiology: IJMM. 2011; 301(1):69–78. doi: 10.1016/j.ijmm.2010.08.014 PMID: 20943438.

10. Shannon BA, Garrett KL, Cohen RJ. Links between Propionibacterium acnes and prostate cancer. Future oncology. 2006; 2(2):225–32. doi: 10.2217/14796694.2.2.225 PMID: 16563091.

11. Ayub A, Ashfaq UA, Haque A. HBV induced HCC: major risk factors from genetic to molecular level. BioMed research international. 2013; 2013:810461. doi: 10.1155/2013/810461 PMID: 23991421; PubMed Central PMCID: PMC3749539.

12. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nature genetics. 2012; 44(7):765–9. doi: 10.1038/ng.2295 PMID: 22634754.

13. Saha A, Kaul R, Murakami M, Robertson ES. Tumor viruses and cancer biology: Modulating signaling pathways for therapeutic intervention. Cancer biology & therapy. 2010; 10(10):961–78. PMID: 21084867.

14. Nelson WG, De Marzo AM, DeWeese TL, Isaacs WB. The role of inflammation in the pathogenesis of prostate cancer. The Journal of urology. 2004; 172(5 Pt 2):S6–11; discussion S-2. PMID: 15535435.

15. Klein EA, Silverman R. Inflammation, infection, and prostate cancer. Current opinion in urology. 2008; 18(3):315–9. doi: 10.1097/MOU.0b013e3282f9b3b7 PMID: 18382242.

16. Pang KR, Wu JJ, Huang DB, Tyring SK, Baron S. Biological and clinical basis for molecular studies of interferons. Methods in molecular medicine. 2005; 116:1–23. doi: 10.1385/1-59259-939-7:1 PMID: 16007741.

17. Das D, Wojno K, Imperiale MJ. BK virus as a cofactor in the etiology of prostate cancer in its early stages. Journal of virology. 2008; 82(6):2705–14. doi: 10.1128/JVI.02461-07 PMID: 18160432

18. Ahsan N, Shah KV. Polyomaviruses and human diseases. Advances in experimental medicine and biology. 2006; 577:1–18. doi: 10.1007/0-387-32957-9_1 PMID: 16626024.

19. Samanta M, Harkins L, Klemm K, Britt WJ, Cobbs CS. High prevalence of human cytomegalovirus in prostatic intraepithelial neoplasia and prostatic carcinoma. The Journal of urology. 2003; 170(3): 998–1002. doi: 10.1097/01.ju.0000080263.46164.97 PMID: 12913758.

20. Zambrano A, Kalantari M, Simoneau A, Jensen JL, Villarreal LP. Detection of human polyomaviruses and papillomaviruses in prostatic tissue reveals the prostate as a habitat for multiple viral infections. Prostate. 2002; 53(4):263–76. doi: 10.1002/pros.10157 PMID: 12430138.

21. Whitaker NJ, Glenn WK, Sahrudin A, Orde MM, Delprado W, Lawson JS. Human papillomavirus and Epstein Barr virus in prostate cancer: Koilocytes indicate potential oncogenic influences of human papillomavirus in prostate cancer. The Prostate. 2013; 73(3):236–41. doi: 10.1002/pros.22562 PMID: 22851253

22. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. The Journal of pathology. 1999; 189(1):12–9. doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F PMID: 10451482

23. Lin Y, Mao Q, Zheng X, Yang K, Chen H, Zhou C, et al. Human papillomavirus 16 or 18 infection and prostate cancer risk: a meta-analysis. Irish journal of medical science. 2011; 180(2):497–503. doi: 10.1007/s11845-011-0692-6 PMID: 21400096

24. Blaheta RA, Weich E, Marian D, Bereiter-Hahn J, Jones J, Jonas D, et al. Human cytomegalovirus infection alters PC3 prostate carcinoma cell adhesion to endothelial cells and extracellular matrix. Neoplasia. 2006; 8(10):807–16. doi: 10.1593/neo.06379 PMID: 17032497; PubMed Central PMCID: PMC1715925.

25. Maeda E, Akahane M, Kiryu S, Kato N, Yoshikawa T, Hayashi N, et al. Spectrum of Epstein-Barr virus-related diseases: a pictorial review. Japanese journal of radiology. 2009; 27(1):4–19. doi: 10.1007/s11604-008-0291-2 PMID: 19373526.

26. Urisman A, Molinaro RJ, Fischer N, Plummer SJ, Casey G, Klein EA, et al. Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. PLoS pathogens. 2006; 2(3):e25–e. doi: 10.1371/journal.ppat.0020025 PMID: 16609730

27. Lombardi VC, Ruscetti FW, Das Gupta J, Pfost MA, Hagen KS, Peterson DL, et al. Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome. Science. 2009; 326(5952):585–9. doi: 10.1126/science.1179052 PMID: 19815723.

28. Cingöz O, Paprotka T, Delviks-Frankenberry KA, Wildt S, Hu W-S, Pathak VK, et al. Characterization, mapping, and distribution of the two XMRV parental proviruses. Journal of virology. 2012; 86(1): 328–38. doi: 10.1128/JVI.06022-11 PMID: 22031947

29. Silverman RH, Nguyen C, Weight CJ, Klein EA. The human retrovirus XMRV in prostate cancer and chronic fatigue syndrome. Nature reviews Urology. 2010; 7(7):392–402. doi: 10.1038/nrurol.2010.77 PMID: 20517289

30. Arnold RS, Makarova NV, Osunkoya AO, Suppiah S, Scott TA, Johnson NA, et al. XMRV infection in patients with prostate cancer: novel serologic assay and correlation with PCR and FISH. Urology. 2010; 75(4):755–61. doi: 10.1016/j.urology.2010.01.038 PMID: 20371060

31. Fischer N, Hellwinkel O, Schulz C, Chun FKH, Huland H, Aepfelbacher M, et al. Prevalence of human gammaretrovirus XMRV in sporadic prostate cancer. Journal of Clinical Virology. 2008; 43(3):277–83. doi: 10.1016/j.jcv.2008.04.016 PMID: 18823818

32. Aloia AL, Sfanos KS, Isaacs WB, Zheng Q, Maldarelli F, De Marzo AM, et al. XMRV: a new virus in prostate cancer? Cancer research. 2010; 70(24):10028–33. doi: 10.1158/0008-5472.CAN-10-2837 PMID: 20966126; PubMed Central PMCID: PMC3005136.

33. Knox K, Carrigan D, Simmons G, Teque F, Zhou Y, Hackett J Jr, et al. No evidence of murine-like gammaretroviruses in CFS patients previously identified as XMRV-infected. Science. 2011; 333(6038): 94–7. doi: 10.1126/science.1204963 PMID: 21628393.

34. Katzourakis A, Hué S, Kellam P, Towers GJ. Phylogenetic analysis of murine leukemia virus sequences from longitudinally sampled chronic fatigue syndrome patients suggests PCR contamination rather than viral evolution. Journal of virology. 2011; 85(20):10909–13. doi: 10.1128/JVI.00827-11 PMID: 21849443

35. Sfanos KS, Aloia AL, De Marzo AM, Rein A. XMRV and prostate cancer—a 'final' perspective. Nature reviews Urology. 2012; 9(2):111–8. doi: 10.1038/nrurol.2011.225 PMID: 22231291

36. Menendez-Arias L. Evidence and controversies on the role of XMRV in prostate cancer and chronic fatigue syndrome. Reviews in medical virology. 2011; 21(1):3–17. doi: 10.1002/rmv.673 PMID: 21294212.

37. Selth LA, Roberts MJ, Chow CW, Marshall VR, Doi SA, Vincent AD, et al. Human seminal fluid as a source of prostate cancer-specific microRNA biomarkers. Endocrine-related cancer. 2014; 21(4): L17–21. doi: 10.1530/ERC-14-0234 PMID: 24859988.

38. Kannan K, Wang L, Wang J, Ittmann MM, Li W, Yen L. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(22):9172–7. doi: 10.1073/pnas.1100489108 PMID: 21571633; PubMed Central PMCID: PMC3107329.

39. Ren S, Peng Z, Mao J-H, Yu Y, Yin C, Gao X, et al. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell research. 2012; 22(5):806–21. doi: 10.1038/cr.2012.30 PMID: 22349460

40. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PloS one. 2012; 7(2):e30619–e. doi: 10.1371/journal.pone.0030619 PMID: 22312429

41. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10(3):R25–R. doi: 10.1186/gb-2009-10-3-r25 PMID: 19261174

42. Kent WJ. BLAT—the BLAST-like alignment tool. Genome research. 2002; 12(4):656–64. doi: 10.1101/gr.229202. Article published online before March 2002. PMID: 11932250

43. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 2010; 28(5):511–5. doi: 10.1038/nbt.1621 PMID: 20436464; PubMed Central PMCID: PMC3146043.

44. Brooks MJ, Rajasimha HK, Roger JE, Swaroop A. Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes. Molecular vision. 2011; 17:3034–54. PMID: 22162623; PubMed Central PMCID: PMC3233386.

45. Holland C, Mak TN, Zimny-Arndt U, Schmid M, Meyer TF, Jungblut PR, et al. Proteomic identification of secreted proteins of Propionibacterium acnes. BMC microbiology. 2010; 10:230. doi: 10.1186/1471-2180-10-230 PMID: 20799957; PubMed Central PMCID: PMC3224659.

46. Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiology and molecular biology reviews: MMBR. 2008; 72(2):317–64, table of contents. doi: 10.1128/MMBR.00031-07 PMID: 18535149; PubMed Central PMCID: PMC2415747.

47. Rui S, Tse-Dinh YC. Topoisomerase function during bacterial responses to environmental challenge. Frontiers in bioscience: a journal and virtual library. 2003; 8:d256–63. PMID: 12456368.

48. Fitz-Gibbon S, Tomida S, Chiu BH, Nguyen L, Du C, Liu M, et al. Propionibacterium acnes strain populations in the human skin microbiome associated with acne. The Journal of investigative dermatology. 2013; 133(9):2152–60. doi: 10.1038/jid.2013.21 PMID: 23337890; PubMed Central PMCID: PMC3745799.

49. Mak TN, Sfanos KS, Bruggemann H. Draft Genome Sequences of Two Strains of Propionibacterium acnes Isolated from Radical Prostatectomy Specimens. Genome announcements. 2013; 1(6). doi: 10.1128/genomeA.01071-13 PMID: 24356837; PubMed Central PMCID: PMC3868861.

50. McDowell A, Perry AL, Lambert PA, Patrick S. A new phylogenetic group of Propionibacterium acnes. Journal of medical microbiology. 2008; 57(Pt 2):218–24. doi: 10.1099/jmm.0.47489-0 PMID: 18201989.

51. McDowell A, Valanne S, Ramage G, Tunney MM, Glenn JV, McLorinan GC, et al. Propionibacterium acnes types I and II represent phylogenetically distinct groups. Journal of clinical microbiology. 2005; 43(1):326–34. doi: 10.1128/JCM.43.1.326-334.2005 PMID: 15634990; PubMed Central PMCID: PMC540145.

52. Bergh Drott J. The role of microorganisms in prostate cancer development. Umeå Universitet; 2012.

53. Thompson MP, Kurzrock R. Epstein-Barr virus and cancer. Clinical cancer research: an official journal of the American Association for Cancer Research. 2004; 10(3):803–21. PMID: 14871955.

54. Alexeyev OA, Marklund I, Shannon B, Golovleva I, Olsson J, Andersson C, et al. Direct visualization of Propionibacterium acnes in prostate tissue by multicolor fluorescent in situ hybridization assay. Journal of clinical microbiology. 2007; 45(11):3721–8. doi: 10.1128/JCM.01543-07 PMID: 17881550; PubMed Central PMCID: PMC2168516.

55. Webster GF, Leyden JJ. Characterization of serum-independent polymorphonuclear leukocyte chemotactic factors produced by Propionibacterium acnes. Inflammation. 1980; 4(3):261–9. PMID: 7429606.

56. Fischer N, Mak TN, Shinohara DB, Sfanos KS, Meyer TF, Bruggemann H. Deciphering the intracellular fate of Propionibacterium acnes in macrophages. BioMed research international. 2013; 2013:603046. doi: 10.1155/2013/603046 PMID: 23862148; PubMed Central PMCID: PMC3687600.

57. Kim J, Ochoa MT, Krutzik SR, Takeuchi O, Uematsu S, Legaspi AJ, et al. Activation of toll-like receptor 2 in acne triggers inflammatory cytokine responses. Journal of immunology. 2002; 169(3):1535–41. PMID: 12133981.

58. Graham GM, Farrar MD, Cruse-Sawyer JE, Holland KT, Ingham E. Proinflammatory cytokine production by human keratinocytes stimulated with Propionibacterium acnes and P. acnes GroEL. The British journal of dermatology. 2004; 150(3):421–8. doi: 10.1046/j.1365-2133.2004.05762.x PMID: 15030323.

59. Nagy I, Pivarcsi A, Koreck A, Szell M, Urban E, Kemeny L. Distinct strains of Propionibacterium acnes induce selective human beta-defensin-2 and interleukin-8 expression in human keratinocytes through toll-like receptors. The Journal of investigative dermatology. 2005; 124(5):931–8. doi: 10.1111/j.0022-202X.2005.23705.x PMID: 15854033.

60. Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC cancer. 2005; 5:45. doi: 10.1186/1471-2407-5-45 PMID: 15892885; PubMed Central PMCID: PMC1173092.

61. Khoury JD, Tannir NM, Williams MD, Chen Y, Yao H, Zhang J, et al. Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. J Virol. 2013; 87(16): 8916–26. doi: 10.1128/JVI.00340-13 PMID: 23740984; PubMed Central PMCID: PMC3754044.

62. Tang KW, Alaei-Mahabadi B, Samuelsson T, Lindh M, Larsson E. The landscape of viral expression and host gene fusion and adaptation in human cancer. Nature communications. 2013; 4:2513. doi: 10.1038/ncomms3513 PMID: 24085110; PubMed Central PMCID: PMC3806554.

63. Zambrano A, Kalantari M, Simoneau A, Jensen JL, Villarreal LP. Detection of human polyomaviruses and papillomaviruses in prostatic tissue reveals the prostate as a habitat for multiple viral infections. The Prostate. 2002; 53(4):263–76. doi: 10.1002/pros.10157 PMID: 12430138

64. Simmons G, Glynn SA, Komaroff AL, Mikovits JA, Tobler LH, Hackett J Jr, et al. Failure to confirm XMRV/MLVs in the blood of patients with chronic fatigue syndrome: a multi-laboratory study. Science. 2011; 334(6057):814–7. doi: 10.1126/science.1213841 PMID: 21940862; PubMed Central PMCID: PMC3299483.

65. Mendoza R, Silverman RH, Klein EA, Miller AD. No biological evidence of XMRV in blood or prostatic fluid from prostate cancer patients. PloS ONE. 2012; 7(5):e36073. doi: 10.1371/journal.pone.0036073 PMID: 22615749

66. Groom HC, Warren AY, Neal DE, Bishop KN. No evidence for infection of UK prostate cancer patients with XMRV, BK virus, Trichomonas vaginalis or human papilloma viruses. PloS one. 2012; 7(3): e34221. doi: 10.1371/journal.pone.0034221 PMID: 22470540