Common Mechanisms of Virus-Mediated Oncogenesis in Bladder Cancers Arising In Solid Organ Transplant Recipients

Authors

Gabriel J Starrett[1], Kelly Yu[2], Yelena Golubeva[3], Petra Lenz[3], Mary L Piaskowski[1], David Peterson[1], Michael Dean[2], Ajay Israni[4], Brenda Y Hernandez[5], Thomas C Tucker[6], Iona Cheng[7], Lou Gonsalves[8], Cyllene R Morris[9], Shehnaz K Hussain[10], Charles F Lynch[11], Reuben S Harris[12], Ludmila Prokunina-Olsson[2], Paul Meltzer[1], Christopher B Buck[1]*, and Eric A Engels[2]*.

Affiliations

[1] CCR, NCI, NIH

[2] DCEG, NCI, NIH

[3] Leidos Biomedical Research Inc., Frederick, MD, USA

[4] Department of Medicine, Nephrology Division, Hennepin Healthcare System, University of Minnesota-Twin Cities

[5] University of Hawai'i Cancer Center

[6] The Kentucky Cancer Registry, University of Kentucky, Markey Cancer Center

[7] University of California-San Francisco, Department of Epidemiology & Biostatistics, and Helen Diller Family Comprehensive Cancer Center, University of California San Francisco

[8] Connecticut Tumor Registry, Connecticut Department of Public Health

[9] California Cancer Reporting and Epidemiologic Surveillance Program, UC Davis Comprehensive Cancer Center/UC Davis Health

[10] Cedars-Sinai Cancer and Department of Medicine, Cedars-Sinai Medical Center

[11] The Iowa Cancer Registry, The University of Iowa

[12] Howard Hughs Medical Institute, University of Minnesota-Twin Cities

*These authors contributed equally

Abstract

A small percentage of bladder cancers in the general population have been found to harbor polyomaviruses and papillomaviruses. In contrast, up to 25% of tumors of solid organ transplant recipients, who are at an increased risk of developing bladder cancer, have been reported to harbor BK polyomavirus (BKPyV). To better understand the biology of the tumors and the mechanisms of carcinogenesis from any potential oncovirus, we performed whole genome sequencing and transcriptomic sequencing on bladder cancer specimens from 43 transplant patients. Roughly a third of the tumors from this patient population contained viral sequences, among which the most common were from BKPyV (N=9, 21%), JC polyomavirus (N=7, 16%), carcinogenic human papillomaviruses (N=3, 7%), and torque teno viruses (N=5, 12%). BKPyV presence was validated by Large T antigen staining and revealed heterogeneous staining patterns between cases ranging from less than one percent to 100% of tumor tissue staining positive. In most cases of BKPyV-positive tumors, the viral genome was integrated into the host chromosome consistent with microhomology-mediated end joining and, in some cases, resulting in focal amplifications of the tumor genome. In BKPyV-positive tumors, significant changes in the expression of E2F- and DREAM-regulated genes amongst others were also observed consistent with the depletion of RB-family members by BKPyV Large T antigen in the tumors. We identified four mutation signatures in our cases with APOBEC3 and SBS5 mutations being the most abundant. We also identified a widespread mutation signature most similar to that caused by the antiviral, ganciclovir, and several cases harbored a high mutation burden associated with aristolochic acid, a nephrotoxic compound found in some herbal medicines. The results indicate that viruses may play a causal role in the development of bladder cancer among immunosuppressed individuals.

Introduction

At least 20% of all cancers are attributable to viral, bacterial, or parasitic infections (1). The advent of high-throughput deep sequencing has provided unprecedented opportunities to learn how infectious agents are involved in cancer in an unbiased manner. Several previous studies have searched for microbial nucleotide sequences in The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC), and PanCancer Analysis of Whole Genomes (PCAWG) datasets (2, 3). In addition to confirming known associations, such as the presence of human papillomaviruses (HPVs) in cervical cancer, these studies also uncovered rare cases in which viral sequences were unexpectedly found in other major cancers affecting the general population (3).

Despite the immense amount of tumor sequencing data generated to date, the identification of microorganisms in common cancers through these studies has been limited. A more focused assessment of groups at increased risk for virus-associated cancers may be needed. In particular, oncogenic viruses may contribute to a larger fraction of cancer cases among immunosuppressed individuals, such as those with human immunodeficiency virus (HIV) infection and organ transplant recipients.

Roughly a dozen “high-risk” HPV types cause nearly all cervical cancers, a large majority of other anogenital cancers, and about half of all oropharyngeal cancers (4). The carcinogenic effects of these small circular double-stranded DNA viruses are primarily dependent on expression of the E6 and E7 oncogenes which, among a wide range of other functions, inactivate the tumor suppressor proteins p53 and Rb, respectively (5-9).

Polyomaviruses share many biological features with papillomaviruses. In particular, polyomavirus T antigens perform many of the same functions as papillomavirus oncoproteins and are similarly oncogenic in cellular and animal models (10). Merkel cell polyomavirus (MCPyV) has been identified as an etiological factor in a rare skin cancer, Merkel cell carcinoma (11-13). Another human polyomavirus, BK polyomavirus (BKPyV), has a long-debated history as a candidate cancer-causing virus. Several case reports have described the detection of BKPyV in bladder cancers arising in transplant recipients, and kidney recipients who develop BKPyV viremia or BKPyV-induced nephropathy (BKVN) after transplant are at increased risk of bladder cancer (14-17).

Similar to HPV-induced cervical and oropharyngeal cancers, bladder cancers exhibit somatic point mutations that are largely attributable to the activity of APOBEC3 family cytosine deaminases (18-20). These enzymes normally function as innate immune defenses against viruses by deaminating cytosines in single-stranded DNA, leading to hypermutation of the viral genome (21). Commonly, APOBEC3 enzymes, particularly APOBEC3A and APOBEC3B, can become dysregulated and cause carcinogenic damage to cellular DNA during the development of various types of cancer (22). APOBEC3A and APOBEC3B are upregulated in response to expression of HPV E6 and E7, and APOBEC3A can restrict HPV replication (23-27). The large T antigens (LTAgs) of BKPyV and JC polyomavirus (JCPyV, a close relative of BKPyV) also upregulate APOBEC3B expression and activity (28-30).

To test the hypothesis that oncogenic viruses are etiologically important for bladder cancers arising in immunosuppressed individuals, we evaluated archived tissues from 43 solid organ transplant recipients who developed this malignancy. We performed total RNA sequencing and whole genome sequencing (WGS) from these tissues. We utilized high-sensitivity methods and comprehensive reference sequences for conserved viral proteins to identify known viral species

and to search for divergent viruses. Once viruses were identified, we further evaluated the sequence data for integration events, point mutations, mutation signatures, and differentially expressed genes to identify differences correlating with the presence of these viruses and their integration state.

Materials and Methods

Sample acquisition and ethics

The Transplant Cancer Match (TCM) Study is a linkage of the US national solid organ transplant registry with multiple central cancer registries (https://transplantmatch.cancer.gov/). We used data from this linkage to identify cases of in situ or invasive bladder cancer diagnosed among transplant recipients. Staff at five participating cancer registries (California, Connecticut, Hawaii, Iowa, Kentucky) worked with hospitals in their catchment areas to retrieve archived pathology materials for selected cases.

We obtained twenty 10-micron sections from formal-fixed paraffine-embedded (FFPE) blocks for cases with available material. At each originating institution, the microtome blade was cleaned with nuclease-free water and ethanol between samples. Single 5-micron sections leading and trailing the twenty sections used for nucleic acid isolation were saved for histochemistry.

The TCM Study is considered non-human subjects research at the National Institutes of Health, because researchers do not receive identifying information on patients, and the present project utilizes materials collected previously for clinical purposes. The TCM Study was reviewed, as required, by human subjects committees at participating cancer registries.

Nucleic acid isolation

Samples were simultaneously deparaffinized and digested using 400 µL molecular-grade mineral oil (Millipore-Sigma) and 255 µL Buffer ATL (Qiagen) supplemented with 45 µL of proteinase K (Qiagen). Samples were incubated overnight at 65°C in a shaking heat block. Samples were spun at 16,000 x g in a tabletop microcentrifuge for one minute to separate the organic and aqueous phases. Depending on the presence of visible remaining tissue, some samples were subjected to one or two additional two-hour long digests by the addition of 25 µL of fresh proteinase K buffer. Lysates were stored at 4°C until RNA or DNA isolation.

Lysates were spun at 16,000 x g in a tabletop microcentrifuge for one minute. For DNA isolation, 150 µL of supernatant was moved to a new 1.5 mL tube. 490 µL of binding buffer PM (Qiagen) and 10 µL of 3M sodium acetate were added to the lysate. The mixture was then added to a Qiaquick spin column and spun at 16,000 x g for 30 seconds. Flow-through was reapplied to the spin column for complete binding. The column was washed first with 750 µL of Buffer PE (Qiagen) and then 750 µL of 80% ethanol, spinning at 16,000 x g for 30 seconds and discarding flow-through each time. The column was dried by spinning it at 16,000 x g for 5 minutes. Collection tubes were discarded, and the column was moved to a new microcentrifuge tube. 50 µL of pre-warmed, 65°C 10% buffer EB was applied to the column and incubated for 1 minute. The tube was then spun at 16,000 x g for 2 minutes. DNA quantity and quality were assessed by Qubit (Thermo Fisher) and spectrophotometry (DeNovix). DNA was stored at -20°C until used for library preparation.

For RNA isolation, 150 µL of the remaining clarified lysate was moved to a new tube. 250 µL of buffer PKD (Qiagen) was added and vortexed to mix. The remainder of the RNA extraction process was carried out using a RNeasy FFPE Kit (Qiagen) according to the manufacturer's protocol. RNA quantity and quality were assessed by spectrophotometry (DeNovix) and TapeStation (Agilent).

Immunohistochemistry

FFPE 5-µm thick tissue sections mounted on charged glass slides were stained with antibody against Large T Antigen, clone PAb416 (Sigma Millipore, cat. DP02) which detects LTAg from multiple polyomaviruses. Slides were baked in a laboratory oven at 60□C for 1 hour prior to immunostaining on Ventana Discovery Ultra automated IHC stainer upon following conditions: CC2 (pH9) antigen retrieval for 64 min at 96□C, antibody at concentration 0.5 µg/ml in Agilent antibody diluent (cat. S3022) for 32 min at 36□C, Anti-Mouse HQ-Anti HQ HRP detection system for 12 min with DAB for 4 minutes and Hematoxylin II counterstain for 8 minutes. After washing per manufacturer's instructions, slides were incubated in tap water for 10 min, dehydrated in ethanol, cleared in xylene, coverslipped with Micromount media (Leica Biosystems) and scanned on AT2 slide scanner (Leica Biosystems) for pathology review. FFPE sections of cell pellets transfected with LTAg and commercial slides of SV40 infected tissue (Sigma, cat. 351S) were used as positive controls.

Library preparation and sequencing

50-250 ng of isolated DNA was fragmented in microtube-50 using a Covaris sonicator with the following settings: peak power: 100, duty factor: 30, cycles/burst: 1000, time: 108 seconds. End-repair and A-tailing were performed on fragmented DNA using the KAPA HyperPrep Kit (Roche). NEB/Illumina adaptors were ligated onto fragments with KAPA T4 DNA Ligase for 2 hours at 20°C then treated with 4 µL USER enzyme (NEB) for 15 minutes at 37°C to digest uracil-containing fragments. Ligation reactions were cleaned up using 0.8x AMPure XP beads using the KAPA protocol. NEB dual-index oligos were added to the adaptor-ligated fragments and amplified for 6-8 cycles (depending on the amount of input fragmented DNA) using KAPA HiFi HotStart ReadyMix (Roche). Final amplified libraries were cleaned using 1x AMPure beads with the recommended KAPA protocol. Ribosomal sequence-depleted cDNA libraries were

prepared using 50 ng of total RNA with the SMARTer Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian (Takara) following the manufacturer instructions for FFPE tissues. Final RNA and DNA libraries were assessed for size and quantity by Agilent TapeStation

DNA libraries were sequenced on the Illumina NovaSeq 6000 at the Center for Cancer Research (CCR) Sequencing Facility. RNA libraries were sequenced on the Illumina NovaSeq 6000 and NextSeq 550 in high output mode at the CCR Genomics Core. Sequencing metrics are reported in Supplemental Table S1.

Sequence alignments

Reads were trimmed using Trim Galore 0.6.0 with default settings. RNAseq reads were initially aligned using STAR aligner 2.5.3ab (31) against a fusion reference human genome containing hg38, all human viruses represented in RefSeq as of December 2018 (Supplemental Table S2), and all papillomavirus genomes from PaVE https://pave.niaid.nih.gov (32). Default parameters were used with the following exceptions: chimSegmentMin=50, outFilterMultimapNmax=1200, outFilterMismatchNmax=30, outFilterMismatchNoverLmax=1. Any reads that had less than 30 bp of perfect identity were excluded. Trimmed DNA reads were aligned with Bowtie2 (2.3.4.3) using the --very-sensitive setting to the same reference genome as mentioned above excluding RNA viruses (33). Alignments were sorted and duplicate sequences were flagged using Picard 2.20.5. Indel realignments and base quality recalculations were conducted using GATK.

Virus detection and integration analysis

All WGS reads not mapping to the human genome were *de novo* assembled using MEGAHIT (1.1.4) with default parameters (34). All trimmed RNA reads were assembled using RNASPAdes (35). Assembled contigs were annotated using BLASTn and BLASTx against the NCBI nt database (October 2021) for closely related species, and CenoteTaker2

(https://github.com/mtisza1/Cenote-Taker2) was used to identify more divergent species in contigs ≥1000bp (36). Depth and breadth of coverage of viral species were normalized by total number of human reads and length of the viral genome. Only species with ≥10% genome coverage and a normalized depth ≥10 in a given sample were considered as hits. Viral read k-mers were cross-compared against samples for uniqueness to identify index hopping or potential contamination between samples. Rearrangements in the BKPyV regulatory region were analyzed and annotated using BKTyper (37). Oncovirus Tools (github.com/gstarrett/oncovirus_tools) was used to isolate and assemble integration junctions (38, 39).

Transcriptome clustering and differential gene expression analysis

Counts from STAR were input into R and normalized using the DESeq2 vst function (40). Batch effects were removed using the R package limma and the function RemoveBatchEffects. These normalized counts were input into the R package ConsensusClusterPlus. Pathway analysis was conducted using Enrichr (https://amp.pharm.mssm.edu/Enrichr) (41, 42).

Somatic point mutation, structural variant, and copy number variant calling

Point mutations were called using Mutect2, VarScan2 and lofreq with default parameters (43, 44). Consensus calls between these variant callers were performed using SomaticSeq (3.3.0) (45). Likely germline variants were annotated and removed using SnpSift and dbSNP v152. Likely somatic point mutations were further filtered by the following criteria: SomatiqSeq PASS filter, ≥10% allele frequency, ≥4 reads supporting the variant allele, and ≥8 reads of total coverage of that position. Common mutations in cancer were annotated using SnpSift and COSMIC. Structural variants were called using BreakDancer (46) and normalized with a custom Python script. Copy number variants in tumor WGS datasets were called using GATK4 CNV to compare them to a panel comprised of the normal-tissue WGS datasets generated in this study.

Recurrent copy number variants within polyomavirus-containing tumors or tumors with no virus were determined using GISTIC2 with default parameters. Visualization and variant calling for BKPyV was performed on alignments against a BKPyV genotype Ib-2 isolate (accession number: AB369087.1).

Mutation signature analysis

Mutation signature analysis was conducted using the likely somatic variants passing all the above criteria. Mutational Patterns and Somatic Signatures R packages were used for de novo somatic mutations signature analysis.

Statistical analysis

All graphs were made in the R statistical environment (4.0.3) using the package ggplot2. Survival analysis was conducted using the R packages survminer with Cox regression and the following covariates: age at diagnosis, summary stage, sex, years since transplant, and virus status.

Results

Bladder cancers from transplant recipients

The study population was comprised of 43 U.S. bladder cancer cases from patients who developed cancer after receiving solid organ transplantation (Table 1). Seventy percent were male and 70% non-Hispanic white. The median age at cancer diagnosis was 63 years (range: 27-82). The most commonly transplanted organ was the kidney (56%), followed by heart and/or lung (33%) and liver (9%). Primary tumors were roughly an equal mixture of high- and low-grade carcinomas diagnosed a median of 5.7 years after transplantation. Twelve cases were categorized as in situ as defined by the Surveillance, Epidemiology, and End Results (SEER) Program, with two of those cases being transitional cell carcinomas in situ and ten cases being

noninvasive papillary transitional cell carcinomas. Invasive cases were mostly categorized into the localized stage (n=20, 46%), which includes tumors that have invaded into the mucosa, submucosa, muscle or subserosa. The eleven remaining cases either had regional or distant invasion or metastasis. We successfully generated WGS data for 39 primary tumors, 3 metastases, and 3 normal tissues, with a median of 31x coverage across the human genome (range: 14-55x). We generated total RNA sequencing data for 42 primary tumors, 5 metastases, and 15 normal tissues, with a median of 30 million reads per sample (range: 4-65.5 million).

Viruses in bladder cancers from transplant recipients

Analysis of WGS data for 39 primary tumors identified one or more virus species in 16 specimens (41%) (Figure 1A). RNA sequencing on tumor samples for which WGS data could not be obtained revealed three additional cases containing viral sequences (45% of samples overall). Among the 20 virus-positive primary tumors, the majority harbored BKPyV (n=9) or JCPyV (n=7). High-risk HPV genotypes 16 and 51 were detected in one and two tumors, respectively. A low-risk papillomavirus, HPV28, was observed in TBC33. One BKPyV-containing tumor (case TBC05) also harbored relatively abundant amounts of HPV20, a *Betapapillomavirus* that can cause cutaneous squamous cell carcinoma in animal model systems (47). Only two reads mapped to HPV20 in the RNA dataset for case TBC05, suggesting that the viral genome was transcriptionally silent. Sequencing of metastases confirmed the presence of BKPyV in TBC06, JCPyV in TBC34, and HPV16 in TBC10. Additionally, sequencing of two separate tumor sections for TBC03 and TBC09 confirmed the presence of BKPyV in both.

WGS from TBC16, TBC17, TBC18, TBC19, TBC20, TBC21, TBC22, TBC23, TBC24, TBC27 had low numbers of reads mapping to the BKPyV genome that were judged to be attributable to low levels of index-hopping from TBC01, a papillary urothelial neoplasm of low malignant

potential (PUNLMP) that had extremely high BKPyV coverage and was sequenced in the same run. Considering this, along with the absence of RNA reads supporting the presence of BKPyV, we scored these tumors virus-negative.

A separate set of searches aimed at identifying divergent members of other virus groups revealed that several tumors (TBC08, TBC14, TBC25, TBC28, and TBC35) and normal tissues (TBC35, TBC28, TBC39) harbored torque teno virus (TTV) sequences from either WGS or RNA sequencing (Supplemental Table S3). Epstein Barr virus was most strongly detected in one normal lymph node (TBC23) and, at low levels, in tumors TBC07 and TBC08.

BKPyV sequences detected in this study came from every genotype except IV (Figure 2A). One patient with a BKPyV-positive tumor had a documented history of BKVN. We identified unambiguous BKPyV integration sites in five of the nine BKPyV-positive tumors and in one normal tissue (Figure 2B & Table 2). For three tumors a single integration junction was identified, and in TBC02 three junctions were identified. In case TBC03, two separate sections from separate blocks of the primary tumor were sequenced. In one of the two sections, 11 integration junctions were identified across seven chromosomes. Interestingly, only three of the junctions could be identified in the second section of the tumor, suggesting either that these junctions were not present throughout the tumor or there was insufficient tumor purity/sequencing depth to detect them.

Integration appeared consistent with a microhomology-mediated end-joining (MMEJ) model for integration, as 20 of 25 junctions (80%) had microhomology greater than or equal to 2bp. In this model, which has previously been proposed for both HPV- and MCPyV-associated tumors (39, 48, 49), microhomologies between the virus and host genomes initiate DNA repair processes that can, in some cases, lead to tandem head-to-tail concatemeric repeats of the viral genome

as well as focal amplifications of the flanking host chromosome. The amplification events are thought to occur either through the formation of a circular DNA intermediate or through additional microhomology-mediated break-induced replication events that occur in concert with the initial MMEJ event that initiates integration. In support of this concept, focal amplifications adjacent to BKPyV integration sites were identified in three patient tumors. In TBC03, amplification of a 17 kb region of chromosome 1 flanking a multi-copy BKPyV integrant was observed in two tumor sections (Figure 2C). In TBC04, a 15 kb single-copy amplification of chromosome 3 was identified. Lastly, a 195 kb region of chromosome 6 was found amplified next to the BKPyV integration junction in TBC08. Twenty-two of the identified 25 junctions (88%) intersected protein coding genes and thus might conceivably affect gene expression or function. Lastly, a possible HPV51 integration event in TBC32 appears to have involved a Mer4-in retrotransposon.

BKPyV RNA and DNA abundance by sequencing generally did not correspond to specimen tumor purity or the number of LTAg+ cells (Figure 3A and B). Gene level analysis of the RNA sequencing data revealed that nearly all polyomavirus-positive tumors predominantly expressed the T antigens, with lower expression of the late genes VP1 and VP2 (encoding the major and minor capsid proteins, respectively) (Figure 3A and C, Supplemental Figure S1). The LTAg open reading frames (ORFs) in these cases were truncated before the helicase domainthrough deletions, frameshifts, or point mutations. An exception was the BKPyV-positive PUNLMP case TBC01, which showed balanced expression of both early and late regions.

BKPyV isolates found in cases of polyomavirus nephropathy typically have rearrangements in the regulatory region that enhance viral replication in cell culture. TBC01 was the only polyomavirus-positive tumor showing evidence of regulatory region rearrangements (Supplemental Figure S1). These observations, together with the exceptionally high coverage

depth for BKPyV (Figure 3A), suggest a scenario in which the TBC01 tumor was productively infected with a pathogenic BKPyV variant.

Immunohistochemistry (IHC) for polyomavirus LTAg was performed on 18 specimens suspected to contain polyomaviruses and two negative control specimens determined to be free of detectable viral sequences. Control sections were negative for TAg staining, whereas 11 of the 18 specimens that contained polyomavirus sequences showed at least some evidence for TAg positivity (Figure 3B and Supplemental Figure S1). Three tumors scored as BKPyV sequence-positive had strong to moderate LTAg staining in greater than 80% of tumor cells, but the other BKPyV-positive tumors had more variable staining. Moderate to weak staining was visible in less than 0.5% cells in the primary tumor for TBC06 (Figure 3D), but strong staining was observed in about 25% of cells in the metastasis. For TBC09, one sample of the tumor was >90% positive for LTAg staining and another sample was less than 25% positive (Supplemental Figure S1). Analysis of WGS data for the two samples of TBC09 did not indicate a collision of more than one tumor, so this may represent transcriptional silencing or loss of BKPyV DNA from one part of the tumor. The normal tissue for TBC09 showed BKPyV RNA and DNA coverage along a small portion of the regulatory region and small T antigen, but no staining for LTAg. Although TBC01 had very high levels of BKPyV DNA and RNA reads, it had the lowest observed proportion of LTAg-positive cells (<1% in a section that was >95% tumor). LTAg-positive cells in the TBC01 sample were almost entirely localized to the luminal margin of the tumor (Figure 3D).

In the cases that were positive for JCPyV, DNA and RNA coverage depth was lower than observed for BKPyV-positive tumors, and in several DNA-positive cases JCPyV transcription was not detected (Figure 1). JCPyV reads were detected in three samples from case TBC12 including the primary tumor, tumor-positive lymph node, and adjacent normal bladder wall

(Supplemental Figure S2). IHC detected LTAg in JCPyV-positive case TBC13, but not in any tissue samples for TBC12.

For two of three cases harboring HPV types known to cause cervical cancer (HPV16 and HPV51), transcripts encoding the E6 and E7 oncogenes were detected (Supplemental Figure S3). In one HPV16+ case (TBC10), viral oncogene RNA expression was detected in both the primary and metastatic specimens. Relative to the primary tumor, the metastasis appeared to exhibit a shift toward increased expression of the E6* splice variant commonly observed in advanced cervical cancer (Supplemental Figure S3B). One case harbored DNA sequences aligning to HPV20 and a single case harbored DNA aligning to HPV28. No RNA reads were detected for these cutaneous HPV types.

For the five TTV-positive tumors, the WGS analyses did not show evidence of integration. However, we were unable to assemble complete circular genomes for any of the TTVs. The missing segments all overlapped the GC-rich origin of replication that forms stable hairpins and is therefore relatively resistant to sequencing with standard Illumina technology (50). All observed TTV ORF1 sequences belonged to the *Alphatorquevirus* genus and had 51-100% amino acid identity to previously reported TTV strains (Supplemental Table S3).

Mutation signature analysis

The overall tumor mutational burden did not show a clear correlation with the presence of viral sequences (Figure 4A). We analyzed likely somatic point mutations from all tumors and deconvoluted mutation signatures (Figures 4B and 4C, Supplemental Figure S4). As expected for bladder cancer, we commonly observed single-base substitution 2 (SBS2) and SBS13 (both characteristic of APOBEC3 mutagenesis, N=13 cases) and SBS5 (associated with smoking history and ERCC2 mutations, N=11 cases).

Interestingly, four tumors (TBC16, TBC28, TBC31, TBC33) carried a predominant SBS22 signature, which is caused by the chemical aristolochic acid found in the birthwort family of plants. Cases with this signature showed a very high mutational burden (Figure 4A). In support of the idea that cases with strong SBS22 signatures arose through environmental exposure, one such case, a kidney recipient, was previously diagnosed with Chinese herbal medicine nephropathy. The final deconvoluted signature closely matched the mutation spectrum caused by the deoxy-guanosine analog, ganciclovir, recently identified in hematopoietic stem cell transplant recipients (Figure 4B) (51). Treatment with ganciclovir is common in solid organ transplant recipients to prevent human cytomegalovirus (HCMV)-mediated disease; however, ganciclovir treatment history was unavailable for these cases to confirm that this is the origin of this mutation signature in these cases.

Recurrent somatic mutations

Numerous cellular genes were found to recurrently harbor nonsynonymous, nonsense, and frameshift mutations (Figure 4E). The spectrum of frequently mutated genes is similar to those reported in various types of urothelial carcinoma (e.g., mutations in *KMT2D*, *KDM6A,* and *ARID1A*) (Supplemental Table S4)(18, 52, 53). No nonsynonymous mutations were identified in *FGFR3* or *PIK3CA,* even though these genes are commonly mutated in non-muscle invasive bladder cancer (NMIBC) (54). Mutations in *TP53*, which are common in muscle-invasive bladder cancer (18), were detected in four tumors (Figure 4E).

To address the reproducibility of mutation calls in deep sequencing of FFPE samples, we analyzed the sequences from two independent sections from separate blocks for three tumors (Figure 4D). Comparing the variants called in these tumors, 77-82% of inferred somatic mutations were common to both sections. Furthermore, a similar comparison showed a large

percentage of variants in common between primary tumors and their metastases (Figure 4D). In TBC06, 84% of the likely somatic mutations in the metastasis were found in the primary tumor, whereas only 28% of the likely somatic variants in the primary tumor were found in the metastasis. In an additional primary-metastatic pair (TBC34), we identified a similar proportion of shared “trunk” mutations but the metastasis had more unique, likely somatic variants.

Patterns of copy number changes differed substantially between BKPyV-positive and virus-negative tumors (Figure 5). BKPyV-positive tumors showed moderate enrichments for gains of chromosome segments 1q, 2p, 3p, 5p, 20q and 22q, while losses of chromosome 2q, 5q, 6q, and 10q were also observed more frequently in BKPyV-positive tumors versus virus-negative tumors. Similar differences in copy numbers have been observed for virus-positive and virus-negative Merkel cell carcinoma (39).

Gene expression analysis

Differential gene expression analysis for BKPyV-positive tumors versus virus-negative tumors revealed 1062 genes that were significantly differentially regulated in tumors harboring BKPyV (Figure 6A). Clustering all primary and metastatic tumors by genes with a greater than three-fold difference of expression in the above comparison, we identified three major groups that loosely correspond to the amount of BKPyV DNA and RNA in a tumor (Figure 6C). A notable exception is the BKPyV-positive tumor TBC01, which falls into the cluster mostly containing virus-negative tumors.

The cluster exclusively containing tumors harboring integrated BKPyV is defined by high expression of genes involved in DNA damage responses, cell cycle progression, angiogenesis, chromatin organization, mitotic spindle assembly and chromosome condensation/separation as well as some genes associated with neuronal differentiation. Overall, these tumors have

relatively low expression of keratins and genes associated with cell adhesion. Genes previously shown to be associated with cell proliferation in bladder cancer, such as FGFR3, are not expressed in BKPyV-positive tumors. Notably, tumors harboring BKPyV have significantly higher average *APOBEC3B* expression compared to both normal tissues and tumors not containing any virus (Figure 6B). This observation is maintained after stratifying the cases by the germline variant, rs1014971, known to associate with increased APOBEC3B expression and bladder cancer risk with highest average APOBEC3B expression observed in tumors with both BKPyV and two copies of rs1014971 (Supplemental Figure S5).

In TBC03, the observed BKPyV integration into Breast Cancer Antiestrogen Resistance 3 (BCAR3) results in increased expression. Further evaluation of RNA reads covering this region reveal a general enrichment of sense and antisense reads mapping to positions 93,688,393-93,704,476, corresponding to the amplified chromosomal region observed in the WGS dataset. There is an even greater enrichment of mapped reads between positions 93,694,469-93,696,857. No increases in expression in nearby host genes were observed for other cases and integration events.

Clinical outcome

Multivariate Cox regression revealed that SEER summary stage was the strongest predictor of overall survival with regional or distant stage tumors having shorter overall survival (Regional: HR=17.25, $p=2.45\times10^{-7}$; Distant: HR=23.24, $p=3.41\times10^{-10}$) (Supplemental Figure S6). Age at diagnosis, grade, years since transplantation, and sex were not significant predictors of outcome. Polyomavirus presence had a modest yet statistically significant (HR=2.43, p=0.030) effect indicating that patients whose tumors contained a polyomavirus (BKPyV or JCPyV) had shortened overall survival compared to patients with no detectable virus in their tumors

(Supplemental Figure S6). The number of cases with tumors containing other virus families was too small to make clear inferences about survival.

Discussion

This report presents a comprehensive molecular assessment of 43 bladder cancers arising in solid organ transplant recipients by WGS and total RNA sequencing. DNA and/or RNA sequences of human BK or JC polyomaviruses were detected in 16 tumors (37%). Expression of the polyomavirus LTAg was documented immunohistochemically in ten cases. HPV sequences were detected in six cases, including four cases with HPV types known to cause cervical cancer. Overall, this is a much higher frequency of small DNA tumor virus sequence detection compared to prior surveys of bladder cancers affecting the general population, where fewer than 5% of tumors harbor small DNA tumor virus sequences (3). Polyomavirus detection in NMIBC falls between these two findings, potentially supporting a greater role for virus infection in less advanced bladder cancer (55). The results suggest that human polyomaviruses and papillomaviruses play a causal role in the development of bladder cancer.

BKPyV infection in organotypic urothelial cell culture has been shown to promote cellular proliferation, consistent with the known transforming effects of its T antigens (56). Interestingly, we observed frequent loss of the helicase domain of BKPyV LTAg due to deletions and point mutations. While such deletions are commonly observed in MCPyV-positive Merkel cell carcinoma (11), MCPyV LTAg lacks the p53-inactivating activity of the C-terminal helicase domain of BKPyV. One might thus have expected the C-terminal portion of BKPyV LT to be preserved in tumor cells. We speculate that the loss of the BKPyV helicase domain is driven by other effects on tumor survival (e.g., LTAg might unwind the integrated BKPyV origin of replication and initiate “onion skin” DNA structures leading to chromosomal instability and cell death).

We also observed amplification of the host genome surrounding BKPyV integration sites, consistent with circular DNA intermediates and/or MMEJ break-induced replication, similar to what has been described for HPV and MCPyV-associated tumors (39, 57). While the observed integration sites in this study are unique and have not been observed in Merkel cell carcinoma, HPV16 integration has been reported previously in BCAR3 (58). Elevated expression of BCAR3 has been shown to increase proliferation, motility and invasiveness of estrogen receptor-positive breast cancer cells after treatment with antiestrogens (59, 60).

In five tumors harboring integrated BKPyV sequences (TBC02, TBC03, TBC05, TBC06 and TBC08), we observed significant upregulation of genes associated with cell cycle progression, DNA damage, histones, and the mitotic spindle. Tumors with evidence of BKPyV integration also exhibited significant downregulation of keratins and cell adhesion genes. The latter may contribute to the association of BKPyV tumors being high grade with invasive behavior. BKPyV-positive tumors without evidence of viral integration either showed more moderate gene expression changes or were indistinguishable from BKPyV-negative tumors.

Many of the observed gene expression changes are consistent with known effects of BKPyV infection and the activities of LTAg, which binds Rb-family proteins and alters the active pool of E2F transcription factors in the cell (61, 62). Recent studies have shown that APOBEC3B expression is repressed by the DREAM complex (which is composed of Rb-family proteins and E2F transcription factors (28)) and, accordingly, we found that APOBEC3B is more highly expressed in BKPyV-positive tumors compared to normal tissues and tumors without tumor virus sequences. The mutation signature commonly attributed to APOBEC3B was frequently observed among our cases, but it was not significantly enriched in BKPyV-positive tumors. It is possible that tumors expressing BKPyV LT and increased APOBEC3B manifested greater intratumor heterogeneity, and we were unable to detect these low frequency APOBEC3-mediated variants from FFPE tissue without deeper and more accurate sequencing.

While HPVs are not generally considered causative agents of bladder cancer, they have been detected in rare cases of bladder cancer affecting immunocompetent and immunosuppressed patients (2, 3, 63). In the current study, we identified four tumors with carcinogenic *Alphapapillomavirus* sequences (HPV16 or HPV 51). *Alphapapillomaviruses* are believed to cause cancer through sustained expression of their E6 and E7 oncoproteins, which is usually (though not always) associated with integration of the papillomavirus genome into the tumor genome.

One case in the panel carried sequences of HPV20, a *Betapapillomavirus*. The possible involvement of *Betapapillomaviruses* in skin cancer in the general population remains controversial (64). In epidermodysplasia verruciformis, a rare syndrome caused by defects in zinc-binding proteins EVER1 and EVER2, patients frequently develop non-melanoma skin cancers containing *Betapapillomaviruses* (65). Expression of E6 and E7 from *Betapapillomaviruses* has been shown to promote cell survival in the face of ultraviolet radiation damage and other carcinogenic insults (47, 64, 66). In the context of bladder cancer, it is possible that cutaneous papillomaviruses likewise enable the accumulation of carcinogenic DNA damage. Additionally, identification of HPV28, a low risk *Alphapapillomavirus*, suggests more abundant papillomavirus infections of the bladder than previously assumed, with unknown effects on carcinogenesis.

We identified three bladder cancers in kidney transplant recipients that exhibited abundant mutations attributable to aristolochic acid-mediated DNA adducts. Aristolochic acid is a highly nephrotoxic and mutagenic compound produced by birthwort plants that are used in some types of herbal medicine and that occasionally contaminate grains (67). Exposure to this compound likely contributed to the patients' need for kidney transplantation as well as their eventual development of bladder cancer. Highlighting the highly mutagenic nature of this compound, the three cases with dominant aristolochic acid signature were in the top 4 for total mutation burden (Figure 4). None of the three tumors had detectable oncogenic viral

sequences, but one had detectable TTV. We also identified likely ganciclovir-mediated mutations in most patients indicating that this common treatment to prevent reactivation of HCMV in solid organ transplant recipients may promote mutagenesis in the urinary bladder.

Until recently, this type of molecular assessment from FFPE tissues would have been nearly impossible or badly muddled by the highly damaging effects of formalin fixation and oxidation of nucleic acids over time. Recent advancements in the isolation of nucleic acids, such as low temperature and organic solvent-free deparaffinization, combined with efficient library preparation from low-concentration highly degraded sources, yielded sufficiently high-quality material for WGS variant calling and total RNA seq (68). To address the difficulty of accurately calling somatic variants (which can be problematic even from flash frozen or fresh tissues), we called variants using the consensus of three modern variant callers. The lack of matched normal tissues for most cases is a significant limitation of this work, but our analytical approach accounted for this limitation by focusing the analysis on mutations with >10% allele frequency and those with potential functional effects, and by excluding known germline variants. Our methods were validated internally through the sequencing of separate regions from the same tumor and of primary-metastatic pairs, which reveal similar concordance of mutations as has been reported from flash frozen tissues (69). Our variant calling approach was also validated by the observation that we detected four deconvoluted mutation signatures that match those expected in prior surveys of bladder cancer. However, the low overall coverage of our WGS, remains a limitation of this study.

A potential explanation for the observation that viruses are more prevalent in bladder cancers affecting solid organ transplant recipients compared to cases in the general population is that transplant recipients may often become newly infected through transmission from the donor graft at the time of transplantation, perhaps with a different viral genotype than present in the host previously. This phenomenon is commonly observed in kidney transplantation and is associated with BKVN (70), but has not been documented for heart, lung, or liver transplant

recipients, who are also included in the current study. The lack of detection of BKPyV genotype IV in this study may indicate that this genotype represents a less carcinogenic strain, reminiscent of “low risk” HPV types that are rarely found in tumors. Additionally, this is only the second study to identify JCPyV in bladder tumors, but based on the high degree of similarity between JCPyV and BKPyV, it is not surprising that the two viruses would behave similarly.

Another difference between BKPyV-mediated carcinogenesis and that of HPV and MCPyV may come down to the immunogenicity of their viral oncoproteins. While BKPyV LTAg can provide a growth advantage to cells in culture, the large multi-domain antigen may be relatively immunogenic compared to the much smaller oncoproteins encoded by high-risk HPVs or the highly truncated MCPyV LTAg isoforms typically observed in tumors. Some tumors in this study expressed moderate amounts of late region genes, VP1 and VP2. Expression of these genes could generate virions that would be released from the tumor and recruit immune cells to respond to these foreign antigens. This may also be impacted by the increased expression of APOBEC3B, which can make cells more susceptible to death by exogenous DNA damaging agents and can generate immunogenic neoantigens (71, 72). Several reports of regression of patients’ BKPyV-positive tumors after reduction of immune suppression support the idea that tumors constitutively expressing BKPyV gene products are readily targeted and controlled by the immune system (73-75).

The theoretical immunological costs of viral gene expression for a nascent tumor cell raise the possibility of “hit-and-run” carcinogenesis by tumor viruses in the context of bladder cancer in the general population. Hit-and-run carcinogenesis implies that a virus may play a causal role in early stages of carcinogenesis but then become undetectable at more advanced stages of tumor development. Infection of a premalignant cell may promote its growth and survival through the expression of the viral oncogenes. Additionally, expression of viral oncogenes may also promote genome instability through the expression of the mutagenic APOBEC3 enzymes or other mechanisms that further push the cell towards transformation. In

the general population, BKPyV is observed in upwards of 4% of NIMBC cases, but this falls to less than 0.25% in muscle-invasive bladder cancers (3, 55).

In several tumors we identified heterogeneous expression of LTAg, potentially consistent with a multistage integration and carcinogenesis process proposed by other recent studies on BKPyV-positive urinary tumors from kidney transplant recipients (76, 77). However, our sequencing experiments support a dominant integrated form in most BKPyV-positive tumors in this study. The only exception to this observation is TBC01, which appears to exhibit a viable BKPyV episome with a rearranged regulatory region present in a small subset of tumor cells. This also suggests that archetypal BKPyV, rather than the more pathogenic and replicative rearranged strains found in cases of nephropathy, is more likely to integrate into nascent tumor cells. Another curious observation is that a sample of normal bladder tissue from case TBC09 contained a clonal BKPyV integrant at low frequency in both the RNA and WGS sequencing. This integrant had multiple copies of small T antigen and a large deletion in the regulatory regions (Supplemental Figure S1). RNA sequencing only indicated low levels of small T antigen expression and histology of the section indicates no tumor cells or LTAg staining. This may indicate that integration of BKPyV may happen frequently during infection with no carcinogenic effect, but rare integration events that provide a sufficient growth advantage while not triggering the immune system can drive carcinogenesis.

It remains to be seen whether TTVs contribute to disease in the context of immune suppression. These ubiquitous small single-stranded DNA viruses contain only three open reading frames and no obvious oncogene. A general model is that TTVs establish a chronic infection that the immune system generally keeps in check, but immune suppression results in increases not only in the abundance but also the diversity of TTVs observed in hosts (78). Indeed, detection of TTVs can serve as an indicator of the degree of overall immune suppression in transplant recipients (79). Interestingly, these viruses, like papillomaviruses and

polyomaviruses, also appear to be depleted for APOBEC3 target motifs, consistent with the effects of an evolutionary virus-host arms race (21, 30, 80).

Lastly, we observed the striking result that patients with polyomavirus-positive bladder cancers had significantly shorter overall survival compared to those whose tumors were HPV-positive or virus-negative. While this is a substantial case series, this observation needs to be replicated in further studies to establish its clinical significance. Ever-decreasing sequencing costs will facilitate more of these studies and shed light on rare and understudied tumor types as well as analyses of lower grade and pre-cancerous lesions. The integration and gene expression correlations presented in the current report suggest a causal role for small DNA tumor viruses in bladder cancer, particularly among transplant patients.

Data and code availability

All refseqs for human papillomaviruses were downloaded from PaVE and refseqs for human polyomaviruses were downloaded from NCBI as of November 2018. Our data are available from dbGaP under accession ######. Viral contigs from this study are deposited in GenBank under accessions #######. Code and additional data used in this manuscript are available from www.github.com/gstarrett.

Acknowledgments

This work was funded by the NIH Intramural Research Program and Independent Research Scholar Program. We would like to thank the CCR Genomics Core and CCR Sequencing Facility for their assistance with sequencing. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We would also like to thank the members of the Laboratory of Cellular Oncology for their useful discussion and insights.

We would like to the staff at the Scientific Registry of Transplant Recipients and participating cancer registries who assisted with collection of the data and registry linkages.

Disclaimer: The views expressed in this article are those of the authors and should not be interpreted to reflect the views or policies of the National Cancer Institute, the Health Resources and Services Administration, the Scientific Registry of Transplant Recipients, the cancer registries, or their contractors.

Figure Legends

Figure 1. Detection of viral sequences. A. Primary tumors. B. Metastatic tumors. C. Normal tissues. Viral species are shown on the rows, and each case in the cohort (represented with a TBC number) is a column. Circle size represents the breadth or fraction of the viral genome covered, and color represents the average depth of coverage of the viral k-mers with all coverages over 100 binned together. Specimens without sequencing data have a gray background.

Figure 2. Virus diversity and integration. A. Phylogenetic tree of BKPyV LTAg sequences detected in tumors (red) and reference genotypes. B. Sites of BKPyV integration into host chromosomes are indicated with case numbers. Two separate sections from separate FFPE blocks of the primary tumor were sequenced for case TBC03 (samples TBC03.1 and TBC03.2). Two separate sections were also sequenced for case TBC09, but an integration site was only detected in sample TBC09.1. Integration sites were also detected in normal tissue sample TBC09N. Black and gray bars indicate cytogenetic bands; red bars indicate centromeres. C. Coverage plot of focal amplifications adjacent to BKPyV integration sites in cases TBC03, TBC04, and TBC08. BKPyV integration junctions are indicated by a black arrow. Colored numbers in the body of the graph indicate coverage depth.

Figure 3. BKPyV DNA, RNA and LTAg detection in tumors. A. Barplots showing the abundance of BKPyV DNA and RNA reads standardized to human reads B. Barplots of

histologically estimated percent tumor purity and IHC-positivity for polyomavirus LTAg expression. C. Representative coverage plots for BKPyV DNA (gray) and RNA (red) in BKPyV-positive tumors. Relative copy numbers are indicated by colored boxes and highlight the borders of duplications and deletions in the viral genome. D. Selected images for LTAg IHC highlighting positive staining for BKPyV-positive tumors corresponding to the RNA and DNA coverage plots on the left in C.

Figure 4. Somatic point mutations and mutation signature analysis. A. Tumor mutation burden (TMB, non-synonymous mutations per million bases) for each tumor in this study. Bars are colored by viral positivity (red, BKPyV; green, TTV; blue, HR-HPV; goldenrod, JCPyV) or etiologic agent (aristolochic acid, purple). B. Barplots of the contribution each trinucleotide substitution for the four deconvoluted signatures with the likely mutation process indicated. C. Proportion of each deconvoluted signature that contributes to each sample with virus status indicated by colored circles (red, BKPyV; green, TTV; blue, HR-HPV; goldenrod, JCPyV). D. Number of unique and common trunk mutations in primary-metastatic tumor pairs and tumors with multi-region sequencing. For TBC03, TBC09, and TBC28, branches 1 and 2 refer to two separate areas of the same tumor. For TBC06 and TBC34, branches P and M refer to the primary tumor and metastasis, respectively. E. Oncoprint for the top mutated genes in bladder cancers of transplant patients. Tumors IDs are colored by likely etiology: BKPyV-positive, red; JCPyV-positive, goldenrod; HR-HPV-positive, blue; TTV-positive, green; aristolochic acid, purple; other, black.

Figure 5. Copy number variants. Frequency plots for large copy number variants in BKPyV-positive and virus-negative tumors. Frequency of gains/amplifications are shown in red; losses/deletions are shown in blue.

Figure 6. Differential gene expression in BKPyV-positive tumors. A. Volcano plot of differential gene expression between BKPyV-positive and virus-negative tumors. Significantly differentially expressed genes (q-value < 0.05) with a fold change greater than 2 are in red, genes with a fold change less than 2 are in pink. Non-significant genes are in gray. B. Variance stabilized counts for APOBEC3B from DESeq grouped by normal tissues, virus-negative tumors, and BKPyV-positive tumors showing significantly increased expression in BKPyV-positive tumors . C. Heatmap of Z-scores of significantly differentially expressed genes and genes relevant to bladder cancer grouped by gene ontology. High expression is red, low expression is blue. Tumors are colored by likely etiology: BKPyV-positive, red; JCPyV-positive, goldenrod; HR-HPV-positive, blue; TTV-positive, green; aristolochic acid, purple; other, black. Tumors with evidence of integration are in italics.

Tables:

Table 1. Characteristics of post-transplant bladder cancer cases (N=43)

Characteristic	Statistic	
Sex, N (%)		
Female	13	(30)
Male	30	(70)
Age in years at diagnosis, median (IQR)	63	(59, 70)
Transplanted organ, N (%)		
Kidney	24	(56)
Liver	4	(9)
Heart and/or lung	14	(33)
Pancreas	1	(2)
Race, N (%)		
Non-Hispanic White	30	(70)
Asian/Pacific Islander	7	(16)
Hispanic	5	(12)
Years from transplant to diagnosis, median (IQR)	5.7	(3, 7)
Non-Hispanic White	6.2	(4,7)
Asian/Pacific Islander	5.4	(3,7)
Hispanic	3.9	(1,7)
Summary stage, N (%)		
In situ	12	(28)
Localized	20	(46)
Regional	6	(14)
Distant	5	(12)
Grade, N (%)		
Low	20	(47)
High	21	(49)
Papillary urothelial neoplasm of low malignant potential	1	(2)
Not available	1	(2)

Table 2. BKPyV integrations sites and microhomology

ID	Human sequence match	Virus sequence match	Maximum MH length	MH Sequence	Chromosome	Position	Nearest Gene (Symbol)	Nearest Gene (Ensembl ID)	Distance to Nearest Gene	Nearest RE	Distance to Nearest RE
TBC02	CATCATGATGATGGG	GATGGGCAGCCTA	5	ATGGG	chr2	120378301	INHBB	ENSG00000163083	-26499	MIRb	-45
TBC02	CTCCTGCTCATGAA	CATGAAGGTTAAGCATGCTA	5	ATGAA	chr4	145732354	C4orf51	ENSG00000237136	0	AluSq2	-474
TBC02	ACCATTTAATTCCCAA	AGTGGAAATTAC	2	AC	chr4	145732375	C4orf51	ENSG00000237136	0	AluSq2	-495
TBC03.1	GCCTTTCTTGTGGACTGGGT	ATTTTCATTTCTACTGGGGTCAGGA	0	No overlap	chr1	93693546	BCAR3	ENSG00000137936	0	MIRb	377
TBC03.1	TCTGTTTCTTATTTCAGAA	GGGTTCTCCTGTTTATAAGGTC	2	TC	chr1	93693570	BCAR3	ENSG00000137936	0	MIRb	353
TBC03.1	AGAGCCTTGGTGGTGG	GGTGGCAAACAGTGCAG	5	GGTGG	chr1	93693890	BCAR3	ENSG00000137936	0	MIRb	33
TBC03.1	GATACTTTTTAGACATGC	AACCATGACCTCAGGAAGGA	4	CATG	chr1	93694075	BCAR3	ENSG00000137936	0	MIRb	0
TBC03.1	CCTCAAAGCCACCCACTCC	TTTCCATGAGCCCCAAA	5	CCAAA	chr1	93694843	BCAR3	ENSG00000137936	0	MER5A	-92
TBC03.1	CAATTTTTTTTTTTT	TTTTTTTATTTGTAAGGGTG	7	TTTTTTT	chr12	50449935	LARP4	ENSG00000161813	0	AluSc	0
TBC03.1	TGCAAGGTGCTTCATGTAT	AGGGGGCTTAAAGGATGCA	4	TGCA	chr14	95764390		ENSG00000257275	-6735	MIRb	0
TBC03.1	TAGCCAAAAAAAAAAAGG	AAAAAAAAAGGCCACAG	11	AAAAAAAAAGG	chr20	8525269	PLCB1	ENSG00000182621	0	MamSINE1	154
TBC03.1	CAATTTGGAAAACAAT	ATGCAAGGGCAGTGCACA	2	AT	chr3	73059264	PPP4R2	ENSG00000163605	0	MER103C	69
TBC03.1	TAAAAAGTGTCA	AAGTGTCAATAGAGAAAAA	8	AAGTGTCA	chr4	142307350	INPP4B	ENSG00000109452	0	L2a	0
TBC03.1	TCACACAATTT-TACTCCTCT	ACACTTTTTACACTCCTCTA	8	ACTCCTCT	chr8	140923993	PTK2	ENSG00000169398	0	L2a	0
TBC03.2	GTTGAGTTGGAGCA	CATCTAAATAATCTCTCAAACT	2	CA	chr1	93693160	BCAR3	ENSG00000137936	0	MER5A1	-10
TBC03.2	ACCCAGTCCACAAGAAAGGC	CCAGTAGAAATGAAAAT	0	No overlap	chr1	93693546	BCAR3	ENSG00000137936	0	MIRb	377
TBC03.2	TCTGTTTCTTATTTCAG	GTTCTCCTGTTTATAAGGTC	2	TC	chr1	93693570	BCAR3	ENSG00000137936	0	MIRb	353
TBC04	GAGTGAGTTCATAG	CAACACTGTGGTGAG-TGAGTT	4	GAGT	chr3	5202593	EDEM1	ENSG00000134109	0	L2b	-466
TBC06	CAGACATT-AGGA	TGAGGACCTAACCTGT	4	AGGA	chr2	201676427	MPP4	ENSG00000082126	0	MIR1_Amn	0
TBC08	TCCACTTTCAGTACTT	TGCAAAAAATCAAAT	1	T	chr6	148535326	SASH1	ENSG00000111961	0	AluSq	995
TBC09.1	GGGGCGGTAACTAGAAG	ACTAGAAGCTTGTCGT	8	ACTAGAAG	chr17	61340185	BCAS3	ENSG00000141376	0	L2-3_Crp	0
TBC09N	GAGAAAATAGGACTCGG	AAGATTCGCCTGAGAAAA	7	GAGAAAA	chr18	8169205	PTPRM	ENSG00000173482	0	MER127	-648
TBC09N	TCCATCCTCCTCTAC	CTCCTCTACATTGT	9	CTCCTCTAC	chr3	34028749	LINC01811	ENSG00000226320	130585	L2b	0
TBC09N	ATGTAATATAAAACT	CATGATTTTAACCCAG	0	No overlap	chr3	117678477		ENSG00000239268	0	L2c	0

Supplement:

Figure S1. All SV40 LTAg staining, BKPyV DNA/RNA coverage, and regulatory region structures. A. Representative coverage plots for BKPyV DNA (gray) and RNA (red) in BKPyV-positive tumors. B. Selected images for LTAg IHC highlighting positive staining for BKPyV-positive tumors corresponding to the RNA and DNA coverage plots on the left in A. C. Diagrams of BKPyV NCCR structures and rearrangements in tumors. P=Primary tumor, M=metastatic tumor.

Figure S2. All JCPyV DNA/RNA coverage plots. Representative coverage plots for JCPyV DNA (gray) and RNA (red) in JCPyV-positive tumors.

Figure S3. All HPV DNA/RNA coverage plots. Representative coverage plots for HPV DNA (gray) and RNA (red) in HPV-positive tumors. Diagrams of open reading frames for each respective type are below the coverage plots.

Figure S4. Mutations signature deconvolution. A. Residual sum of squares and explained variance for 2-10 signatures deconvoluted by SomaticSignatures. B. Barplot of base substitution contributions to each of the four deconvoluted signatures from SomaticSignatures. C. Heatmap of cosine similarities of four signatures deconvoluted by Somatic Signatures versus known Single Base Substitution Signatures (SBS). D. NMF rank survey results for 2-10 signature deconvolution by MutationalPatterns. E. Barplot of base substitution contributions to each of the four deconvoluted signatures from MutationalPatterns. F. Heatmap of cosine similarities of four signatures deconvoluted by MutationalPatterns versus known Single Base Substitution Signatures (SBS) with closest matches highlighted in red.

Figure S5. APOBEC3B germline variant and expression by BKPyV status. Stabilized counts of APOBEC3B expression divided by tissue type (primary tumor, normal tissue), BKPyV status (BK), and germline variant rs1014971 status.

Figure S6. Survival and clinical characteristics. Kaplan-meier plots of tumors stratified by SEER summary stage and by BKPyV-postitive tumors versus tumors without detectable viral sequences.

Table S1: Sequencing metrics

Table S2. Reference sequences used in this study

Table S3. Tumor torque teno virus similarities

Table S4. Point mutations and indels

Table S5. Copy number variants

References

1. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. The Lancet Global Health. 2020;8(2):e180-e90.
2. Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. Nat Genet. 2020;52(3):320-30.
3. Cantalupo PG, Katz JP, Pipas JM. Viral sequences in human cancer. Virology. 2018;513:208-16.
4. Graham SV. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. Clin Sci (Lond). 2017;131(17):2201-21.
5. Rosty C, Sheffer M, Tsafrir D, Stransky N, Tsafrir I, Peter M, et al. Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. Oncogene. 2005;24(47):7094-104.
6. Crook T, Tidy JA, Vousden KH. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required. Cell. 1991;67:547-56.
7. Mirabello L, Yeager M, Yu K, Clifford GM, Xiao Y, Zhu B, et al. HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. Cell. 2017;170(6):1164-74.e6.
8. Decaprio Ja. Human papillomavirus type 16 E7 perturbs DREAM to promote cellular proliferation and mitotic gene expression. Oncogene. 2013;33(September):1-3.
9. Barbosa MS, Edmonds C, Fisher C, Schiller JT, Lowy DR, Vousden KH. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. The EMBO journal. 1990;9(1):153-60.
10. Moens U, Macdonald A. Effect of the Large and Small T-Antigens of Human Polyomaviruses on Signaling Pathways. Int J Mol Sci. 2019;20(16).
11. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, et al. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proceedings of the National Academy of Sciences of the United States of America. 2008;105:16272-7.
12. Kassem A, Schöpflin A, Diaz C, Weyers W, Stickeler E, Werner M, et al. Frequent detection of merkel cell polyomavirus in human merkel cell carcinomas and identification of a unique deletion in the VP1 gene. Cancer Research. 2008;68(13):5009-13.
13. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science (New York, NY). 2008;319(5866):1096-100.
14. Gupta G, Kuppachi S, Kalil RS, Buck CB, Lynch CF, Engels EA. Treatment for presumed BK polyomavirus nephropathy and risk of urinary tract cancers among kidney transplant recipients in the United States. Am J Transplant. 2018;18(1):245-52.
15. Vajdic CM, van Leeuwen MT. Cancer incidence and risk factors after solid organ transplantation. Int J Cancer. 2009;125(8):1747-54.
16. Li YJ, Wu HH, Chen CH, Wang HH, Chiang YJ, Hsu HH, et al. High Incidence and Early Onset of Urinary Tract Cancers in Patients with BK Polyomavirus Associated Nephropathy. Viruses. 2021;13(3).
17. Papadimitriou JC, Randhawa P, Rinaldo CH, Drachenberg CB, Alexiev B, Hirsch HH. BK Polyomavirus Infection and Renourinary Tumorigenesis. Am J Transplant. 2016;16(2):398-406.

18. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell. 2017;171(3):540-56 e25.
19. Burns MB, Temiz Na, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nature genetics. 2013;45(9):977-83.
20. Roberts Sa, Lawrence MS, Klimczak LJ, Grimm Sa, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nature genetics. 2013;45(9):970-6.
21. Poulain F, Lejeune N, Willemart K, Gillet NA. Footprint of the host restriction factors APOBEC3 on the genome of human viruses. PLoS Pathog. 2020;16(8):e1008718.
22. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC enzymes: Mutagenic fuel for cancer evolution and heterogeneity. Cancer Discovery. 2015;5(7):704-12.
23. Mori S, Takeuchi T, Ishii Y, Yugawa T, Kiyono T, Nishina H, et al. Human papillomavirus 16 E6 up-regulates APOBEC3B via the TEAD transcription factor. Journal of Virology. 2017:JVI.02413-16.
24. Warren CJ, Westrich JA, Doorslaer KV, Pyeon D. Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression. Viruses. 2017;9(8).
25. Ahasan MM, Wakae K, Wang Z, Kitamura K, Liu G, Koura M, et al. APOBEC3A and 3C decrease human papillomavirus 16 pseudovirion infectivity. Biochemical and biophysical research communications. 2015;457(3):295-9.
26. Warren CJ, Xu T, Guo K, Griffin LM, Westrich Ja, Lee D, et al. APOBEC3A functions as a restriction factor of human papillomavirus. Journal of virology. 2014;89(1):688-702.
27. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz Na, Lorenz LD, et al. Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. mBio. 2014;5(6):1-8.
28. Starrett GJ, Serebrenik AA, Roelofs PA, McCann JL, Verhalen B, Jarvis MC, et al. Polyomavirus T Antigen Induces APOBEC3B Expression Using an LXCXE-Dependent and TP53-Independent Mechanism. MBio. 2019;10(1).
29. Peretti A, Geoghegan EM, Pastrana DV, Smola S, Feld P, Sauter M, et al. Characterization of BK Polyomaviruses from Kidney Transplant Recipients Suggests a Role for APOBEC3 in Driving In-Host Virus Evolution. Cell Host Microbe. 2018;23(5):628-35 e7.
30. Verhalen B, Starrett GJ, Harris RS, Jiang M. Functional Upregulation of the DNA Cytosine Deaminase APOBEC3B by Polyomaviruses. Journal of virology. 2016;90(14):6379-86.
31. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
32. Van Doorslaer K, Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, et al. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res. 2017;45(D1):D499-D506.
33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.
34. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6.

35. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience. 2019;8(9).
36. Tisza MJ, Belford AK, Dominguez-Huerta G, Bolduc B, Buck CB. Cenote-Taker 2 democratizes virus discovery and sequence annotation. Virus Evol. 2021;7(1):veaa100.
37. Marti-Carreras J, Mineeva-Sangwo O, Topalis D, Snoeck R, Andrei G, Maes P. BKTyper: Free Online Tool for Polyoma BK Virus VP1 and NCCR Typing. Viruses. 2020;12(8).
38. Starrett GJ. Oncovirus tools. GitHub2020.
39. Starrett GJ, Thakuria M, Chen T, Marcelus C, Cheng J, Nomburg J, et al. Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma. Genome Med. 2020;12(1):30.
40. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(550):1-21.
41. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7.
42. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.
43. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76.
44. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40(22):11189-201.
45. Fang LT, Afshar PT, Chhibber A, Mohiyuddin M, Fan Y, Mu JC, et al. An ensemble approach to accurately detect somatic mutations using SomaticSeq. Genome Biol. 2015;16:197.
46. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 2009;6(9):677-81.
47. Michel A, Kopp-Schneider A, Zentgraf H, Gruber AD, de Villiers EM. E6/E7 expression of human papillomavirus type 20 (HPV-20) and HPV-27 influences proliferation and differentiation of the skin in UV-irradiated SKH-hr1 transgenic mice. J Virol. 2006;80(22):11153-64.
48. Starrett GJ, Marcelus C, Cantalupo PG, Katz JP, Cheng J, Akagi K, et al. Merkel Cell Polyomavirus Exhibits Dominant Control of the Tumor Genome and Transcriptome in Virus-Associated Merkel Cell Carcinoma. MBio. 2017;8(1).
49. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome research. 2014;24(2):185-99.
50. Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, et al. Discovery of several thousand highly diverse circular DNA viruses. Elife. 2020;9.
51. de Kanter JK, Peci F, Bertrums E, Rosendahl Huber A, van Leeuwen A, van Roosmalen MJ, et al. Antiviral treatment causes a unique mutational signature in cancers of transplantation recipients. Cell Stem Cell. 2021;28(10):1726-39 e6.
52. Nassar AH, Umeton R, Kim J, Lundgren K, Harshman LC, Van Allen EM, et al. Mutational analysis of 472 urothelial carcinoma across grades and anatomic sites. Clin Cancer Res. 2018.

53. Su X, Lu X, Bazai SK, Comperat E, Mouawad R, Yao H, et al. Comprehensive integrative profiling of upper tract urothelial carcinomas. Genome Biol. 2021;22(1):7.
54. Cancer T, Atlas G, Weinstein JN, Akbani R, Broom BM, Wang W, et al. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014;507(7492):315-22.
55. Llewellyn MA, Gordon NS, Abbotts B, James ND, Zeegers MP, Cheng KK, et al. Defining the frequency of human papillomavirus and polyomavirus infection in urothelial bladder tumours. Sci Rep. 2018;8(1):11290.
56. Schneidewind L, Neumann T, Plis A, Bruckmann S, Keiser M, Kruger W, et al. Novel 3D organotypic urothelial cell culture model for identification of new therapeutic approaches in urological infections. J Clin Virol. 2020;124:104283.
57. Czech-Sioli M, Gunther T, Therre M, Spohn M, Indenbirken D, Theiss J, et al. High-resolution analysis of Merkel Cell Polyomavirus in Merkel Cell Carcinoma reveals distinct integration patterns and suggests NHEJ and MMBIR as underlying mechanisms. PLoS Pathog. 2020;16(8):e1008562.
58. Jeannot E, Harle A, Holmes A, Sastre-Garau X. Nuclear factor I X is a recurrent target for HPV16 insertions in anal carcinomas. Genes Chromosomes Cancer. 2018;57(12):638-44.
59. Wallez Y, Riedl SJ, Pasquale EB. Association of the breast cancer antiestrogen resistance protein 1 (BCAR1) and BCAR3 scaffolding proteins in cell signaling and antiestrogen resistance. J Biol Chem. 2014;289(15):10431-44.
60. Wilson AL, Schrecengost RS, Guerrero MS, Thomas KS, Bouton AH. Breast cancer antiestrogen resistance 3 (BCAR3) promotes cell motility by regulating actin cytoskeletal and adhesion remodeling in invasive breast cancer cells. PLoS One. 2013;8(6):e65678.
61. Caller LG, Davies CTR, Antrobus R, Lehner PJ, Weekes MP, Crump CM. Temporal Proteomic Analysis of BK Polyomavirus Infection Reveals Virus-Induced G2 Arrest and Highly Effective Evasion of Innate Immune Sensing. J Virol. 2019;93(16).
62. Harris KF, Christensen JB, Imperiale MJ. BK virus large T antigen: interactions with the retinoblastoma family of tumor suppressor proteins and effects on cellular growth control. Journal of virology. 1996;70(4):2378-86.
63. Guma S, Maglantay R, Lau R, Wieczorek R, Melamed J, Deng FM, et al. Papillary urothelial carcinoma with squamous differentiation in association with human papilloma virus: case report and literature review. Am J Clin Exp Urol. 2016;4(1):12-6.
64. Viarisio D, Muller-Decker K, Accardi R, Robitaille A, Durst M, Beer K, et al. Beta HPV38 oncoproteins act with a hit-and-run mechanism in ultraviolet radiation-induced skin carcinogenesis in mice. PLoS Pathog. 2018;14(1):e1006783.
65. Dell'Oste V, Azzimonti B, De Andrea M, Mondini M, Zavattaro E, Leigheb G, et al. High beta-HPV DNA loads and strong seroreactivity are present in epidermodysplasia verruciformis. J Invest Dermatol. 2009;129(4):1026-34.
66. Viarisio D, Mu ller-Decker K, Zanna P, Kloz U, Aengeneyndt B, Accardi R, et al. Novel ss-HPV49 transgenic mouse model of upper digestive tract cancer2016.
67. Poon SL, Huang MN, Choo Y, McPherson JR, Yu W, Heng HL, et al. Mutation signatures implicate aristolochic acid in bladder cancer development. Genome Med. 2015;7(1):38-.
68. Robbe P, Popitsch N, Knight SJL, Antoniou P, Becq J, He M, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med. 2018;20(10):1196-205.

69. Zhang J, Fujimoto J, Wedge DC, Song X, Seth S, Chow CW, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;346(6206):256-9.
70. Solis M, Velay A, Porcher R, Domingo-Calap P, Soulier E, Joly M, et al. Neutralizing Antibody-Mediated Response and Risk of BK Virus-Associated Nephropathy. J Am Soc Nephrol. 2018;29(1):326-34.
71. Serebrenik AA, Starrett GJ, Leenen S, Jarvis MC, Shaban NM, Salamango DJ, et al. The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. Proc Natl Acad Sci U S A. 2019;116(44):22158-63.
72. Chen Z, Wen W, Bao J, Kuhs KL, Cai Q, Long J, et al. Integrative genomic analyses of APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes, and immunogenicity in multiple cancer types. BMC Med Genomics. 2019;12(1):131.
73. Meier RPH, Muller YD, Dietrich PY, Tille JC, Nikolaev S, Sartori A, et al. Immunologic Clearance of a BK Virus-associated Metastatic Renal Allograft Carcinoma. Transplantation. 2021;105(2):423-9.
74. Cuenca AG, Rosales I, Lee RJ, Wu CL, Colvin R, Feldman AS, et al. Resolution of a High Grade and Metastatic BK Polyomavirus-Associated Urothelial Cell Carcinoma Following Radical Allograft Nephroureterectomy and Immune Checkpoint Treatment: A Case Report. Transplant Proc. 2020.
75. Fu F, Deng W, Yu S, Liu Y, Yu L, Liu R, et al. Occurrence and regression of BK polyomavirus associated carcinoma: a clinical and next-generation sequencing study. Clin Sci (Lond). 2018;132(16):1753-63.
76. Jin Y, Zhou Y, Deng W, Wang Y, Lee RJ, Liu Y, et al. Genome-wide profiling of BK polyomavirus integration in bladder cancer of kidney transplant recipients reveals mechanisms of the integration at the nucleotide level. Oncogene. 2021;40(1):46-54.
77. Wang Y, Liu Y, Deng W, Fu F, Yan S, Yang H, et al. Viral integration in BK polyomavirus-associated urothelial carcinoma in renal transplant recipients: multistage carcinogenesis revealed by next-generation virome capture sequencing. Oncogene. 2020.
78. De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, et al. Temporal response of the human virome to immunosuppression and antiviral therapy. Cell. 2013;155(5):1178-87.
79. Blatter JA, Sweet SC, Conrad C, Danziger-Isakov LA, Faro A, Goldfarb SB, et al. Anellovirus loads are associated with outcomes in pediatric lung transplantation. Pediatr Transplant. 2018;22(1).
80. Warren CJ, Van Doorslaer K, Pandey A, Espinosa JM, Pyeon D. Role of the host restriction factor APOBEC3 on papillomavirus evolution. Virus Evolution. 2015;1(1):vev015-vev.

Figure 1

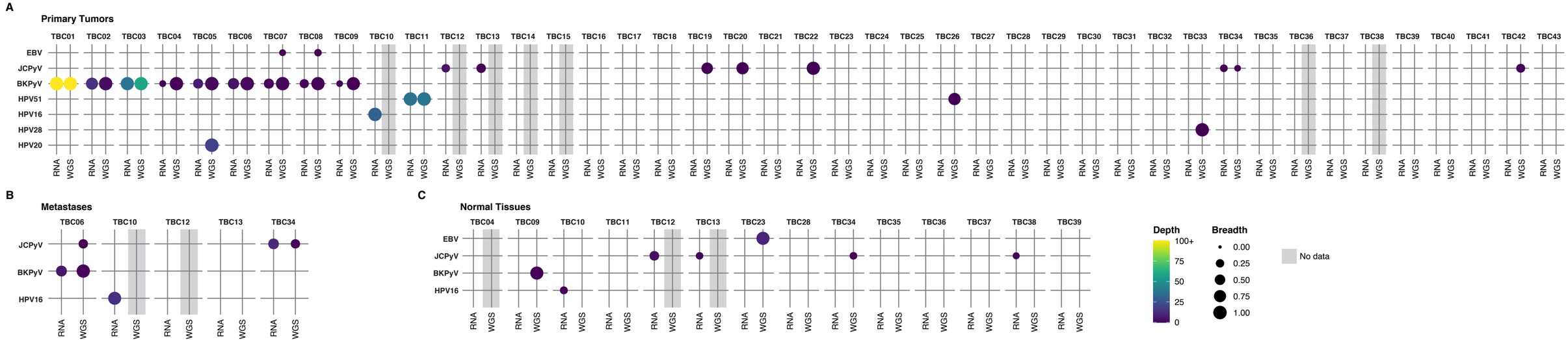

A
Primary Tumors
TBC01 TBC02 TBC03 TBC04 TBC05 TBC06 TBC07 TBC08 TBC09 TBC10 TBC11 TBC12 TBC13 TBC14 TBC15 TBC16 TBC17 TBC18 TBC19 TBC20 TBC21 TBC22 TBC23 TBC24 TBC25 TBC26 TBC27 TBC28 TBC29 TBC30 TBC31 TBC32 TBC33 TBC34 TBC35 TBC36 TBC37 TBC38 TBC39 TBC40 TBC41 TBC42 TBC43
EBV
JCPyV
BKPyV
HPV51
HPV16
HPV28
HPV20
RNA WGS
B
Metastases
TBC06 TBC10 TBC12 TBC13 TBC34
JCPyV
BKPyV
HPV16
RNA WGS
C
Normal Tissues
TBC04 TBC09 TBC10 TBC11 TBC12 TBC13 TBC23 TBC28 TBC34 TBC35 TBC36 TBC37 TBC38 TBC39
EBV
JCPyV
BKPyV
HPV16
RNA WGS
Depth
100+
75
50
25
0
Breadth
0.00
0.25
0.50
0.75
1.00
No data

Figure 2

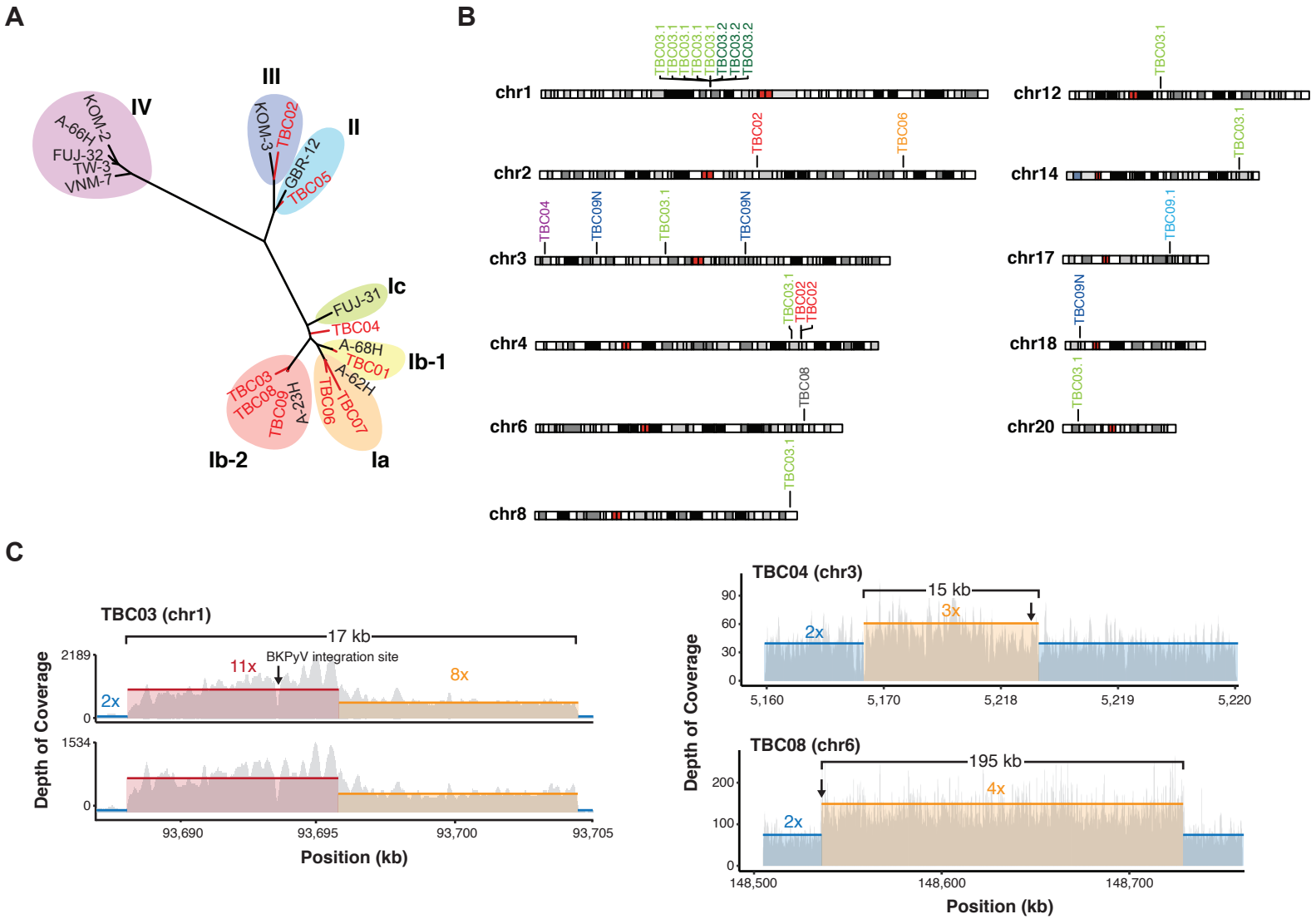

A
IV
III
II
Ic
Ib-1
Ib-2
Ia
KOM-2
A-66H
FUJ-32
TW-3
VNM-7
KOM-3
TBC02
GBR-12
TBC05
FUJ-31
TBC04
A-68H
TBC01
A-62H
TBC03
TBC08
TBC09
A-23H
TBC06
TBC07
B
chr1
chr2
chr3
chr4
chr6
chr8
chr12
chr14
chr17
chr18
chr20
TBC03.1
TBC03.2
TBC02
TBC06
TBC04
TBC09N
TBC09.1
TBC08
C
TBC03 (chr1)
17 kb
BKPyV integration site
11x
8x
2x
2189
0
1534
0
Depth of Coverage
93,690
93,695
93,700
93,705
Position (kb)
TBC04 (chr3)
15 kb
3x
2x
90
60
30
0
5,160
5,170
5,218
5,219
5,220
TBC08 (chr6)
195 kb
4x
2x
200
100
0
148,500
148,600
148,700
Position (kb)

Figure 3

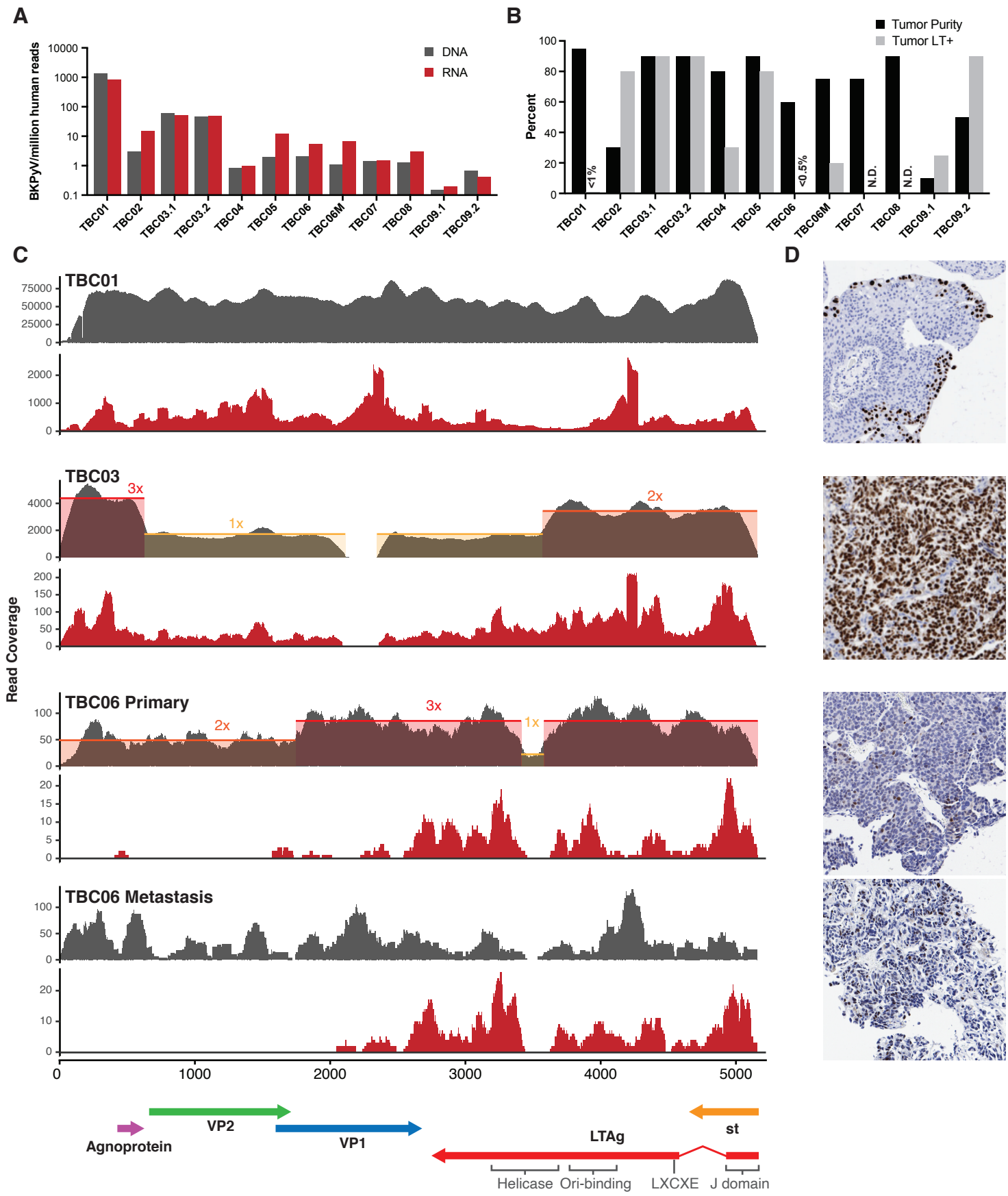

A
BKPyV/million human reads
DNA
RNA
TBC01
TBC02
TBC03.1
TBC03.2
TBC04
TBC05
TBC06
TBC06M
TBC07
TBC08
TBC09.1
TBC09.2
B
Tumor Purity
Tumor LT+
Percent
<1%
<0.5%
N.D.
N.D.
C
D
TBC01
TBC03
3x
1x
2x
Read Coverage
TBC06 Primary
2x
3x
1x
TBC06 Metastasis
Agnoprotein
VP2
VP1
LTAg
st
Helicase
Ori-binding
LXCXE
J domain

Figure 4

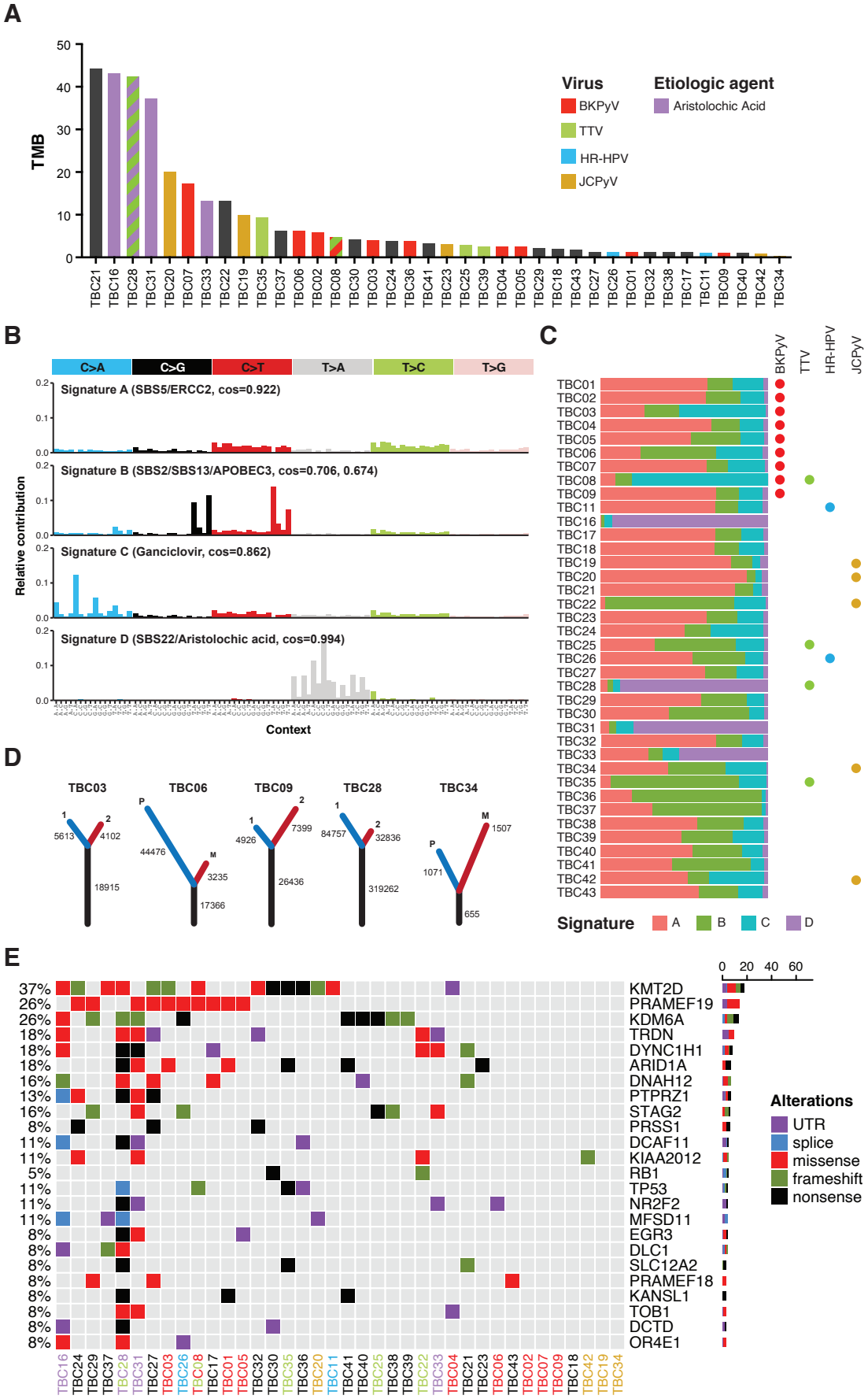

A
TMB
Virus
BKPyV
TTV
HR-HPV
JCPyV
Etiologic agent
Aristolochic Acid
B
C>A
C>G
C>T
T>A
T>C
T>G
Signature A (SBS5/ERCC2, cos=0.922)
Signature B (SBS2/SBS13/APOBEC3, cos=0.706, 0.674)
Signature C (Ganciclovir, cos=0.862)
Signature D (SBS22/Aristolochic acid, cos=0.994)
Relative contribution
Context
C
BKPyV
TTV
HR-HPV
JCPyV
Signature
A
B
C
D
D
TBC03
TBC06
TBC09
TBC28
TBC34
E
Alterations
UTR
splice
missense
frameshift
nonsense
KMT2D
PRAMEF19
KDM6A
TRDN
DYNC1H1
ARID1A
DNAH12
PTPRZ1
STAG2
PRSS1
DCAF11
KIAA2012
RB1
TP53
NR2F2
MFSD11
EGR3
DLC1
SLC12A2
PRAMEF18
KANSL1
TOB1
DCTD
OR4E1

Figure 5

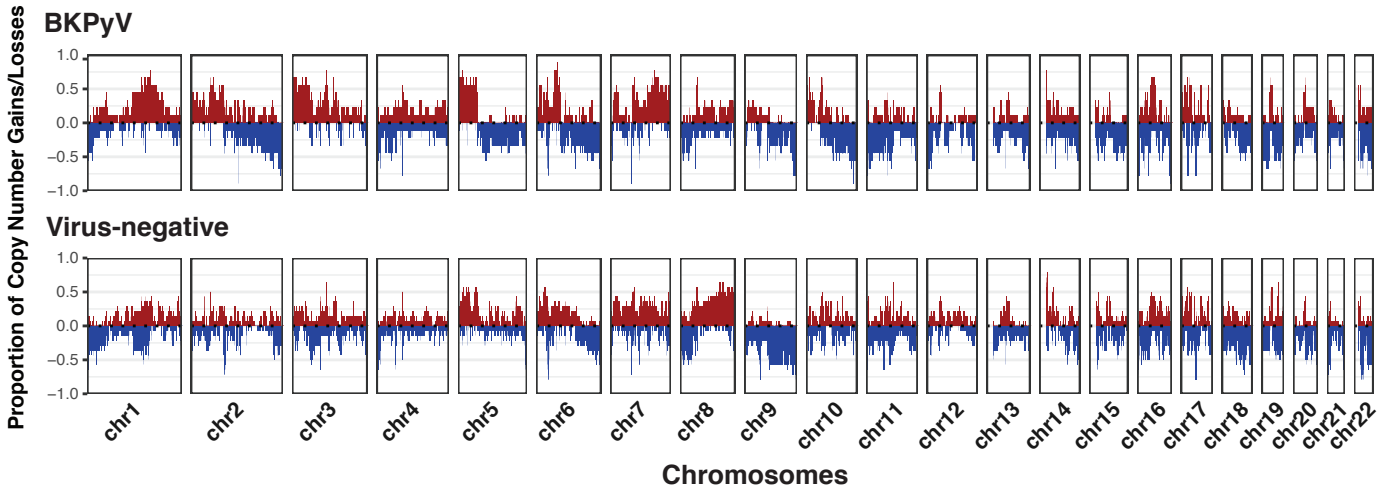
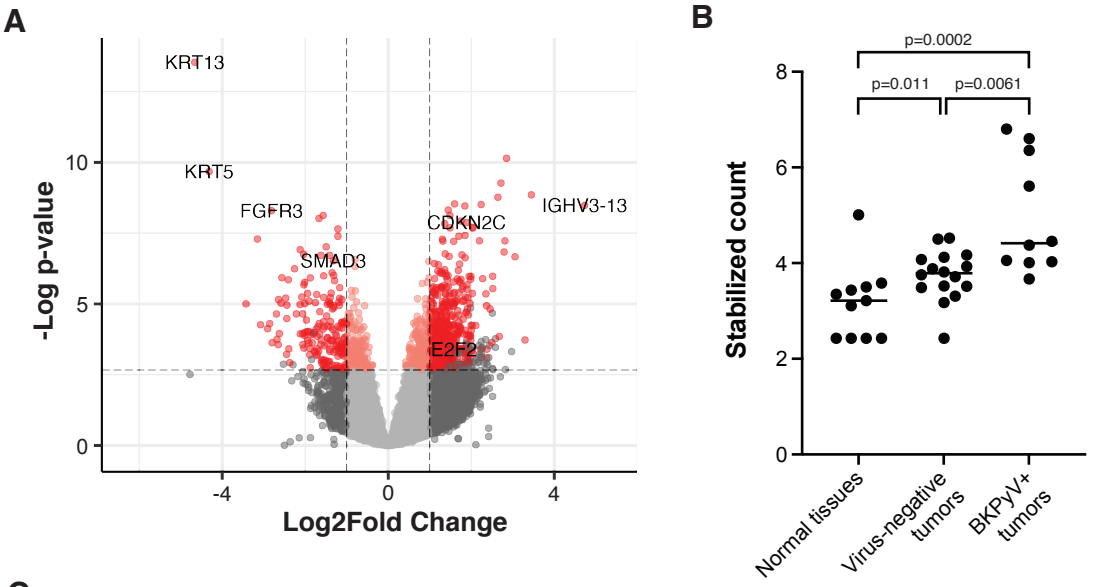
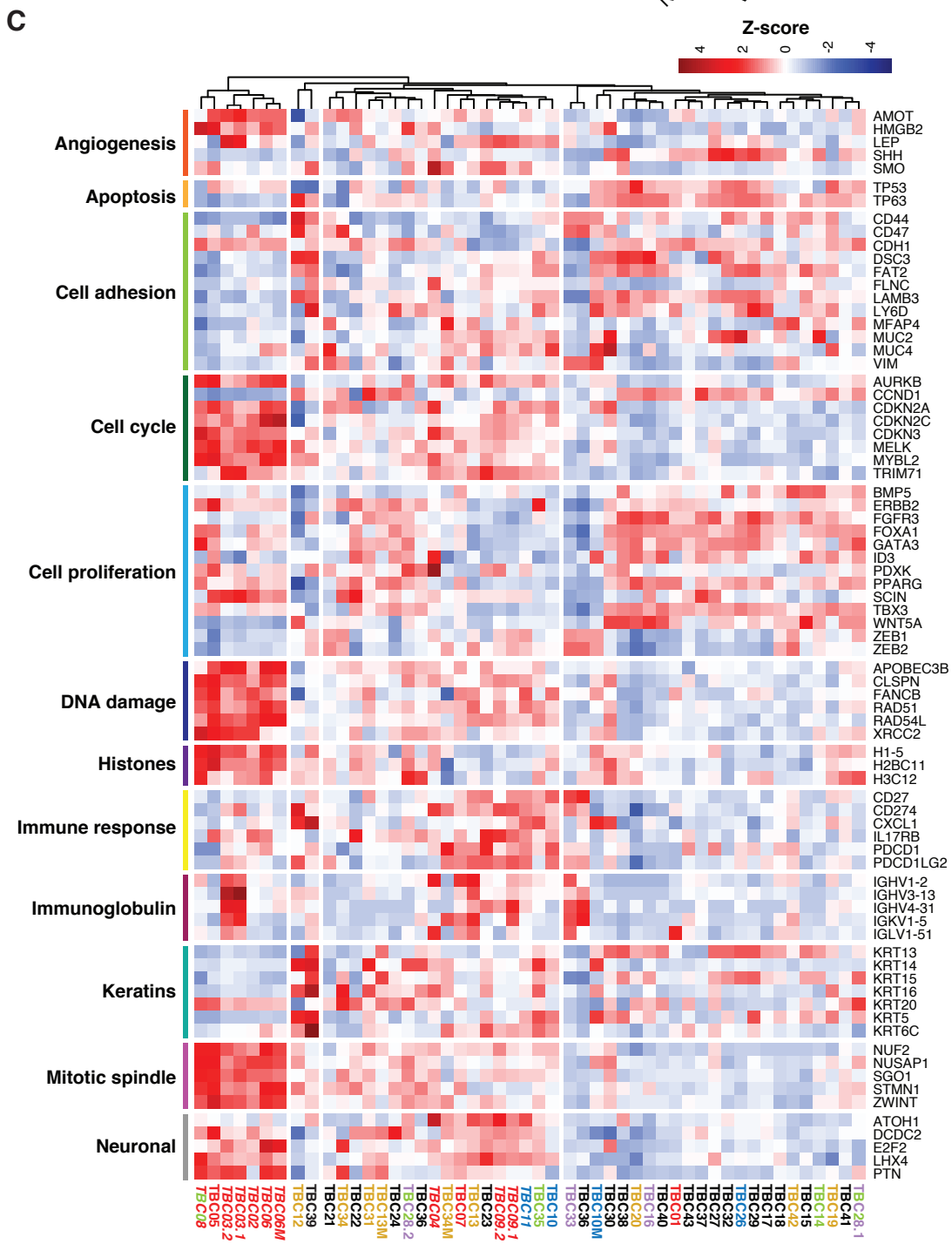

BKPyV
Virus-negative
Proportion of Copy Number Gains/Losses
1.0
0.5
0.0
−0.5
−1.0
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
Chromosomes

Figure 6


A
KRT13
KRT5
FGFR3
SMAD3
CDKN2C
IGHV3-13
E2F2
-Log p-value
Log2Fold Change
B
p=0.0002
p=0.011
p=0.0061
Stabilized count
Normal tissues
Virus-negative tumors
BKPyV+ tumors


C
Z-score
Angiogenesis
Apoptosis
Cell adhesion
Cell cycle
Cell proliferation
DNA damage
Histones
Immune response
Immunoglobulin
Keratins
Mitotic spindle
Neuronal
AMOT
HMGB2
LEP
SHH
SMO
TP53
TP63
CD44
CD47
CDH1
DSC3
FAT2
FLNC
LAMB3
LY6D
MFAP4
MUC2
MUC4
VIM
AURKB
CCND1
CDKN2A
CDKN2C
CDKN3
MELK
MYBL2
TRIM71
BMP5
ERBB2
FGFR3
FOXA1
GATA3
ID3
PDXK
PPARG
SCIN
TBX3
WNT5A
ZEB1
ZEB2
APOBEC3B
CLSPN
FANCB
RAD51
RAD54L
XRCC2
H1-5
H2BC11
H3C12
CD27
CD274
CXCL1
IL17RB
PDCD1
PDCD1LG2
IGHV1-2
IGHV3-13
IGHV4-31
IGKV1-5
IGLV1-51
KRT13
KRT14
KRT15
KRT16
KRT20
KRT5
KRT6C
NUF2
NUSAP1
SGO1
STMN1
ZWINT
ATOH1
DCDC2
E2F2
LHX4
PTN