

# UC Davis

## UC Davis Previously Published Works

### Title

Evidence for virus-mediated oncogenesis in bladder cancers arising in solid organ transplant recipients.

### Permalink

<https://escholarship.org/uc/item/9xd6p2t2>

### Authors

Starrett, Gabriel  
Yu, Kelly  
Golubeva, Yelena  
et al.

### Publication Date

2023-03-24

### DOI

10.7554/eLife.82690

Peer reviewed

# Evidence for virus-mediated oncogenesis in bladder cancers arising in solid organ transplant recipients

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

<sup>1</sup>CCR, NCI, NIH, Bethesda, United States; <sup>2</sup>DCEG, NCI, NIH, Rockville, United States; <sup>3</sup>Leidos Biomedical Research Inc, Frederick, United States; <sup>4</sup>Department of Medicine, Nephrology Division, Hennepin Healthcare System, University of Minnesota, Minneapolis, United States; <sup>5</sup>Cancer Center, University of Hawaii, Honolulu, United States; <sup>6</sup>The Kentucky Cancer Registry, University of Kentucky, Lexington, United States; <sup>7</sup>Department of Epidemiology and Biostatistics, and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, Fremont, United States; <sup>8</sup>Connecticut Tumor Registry, Connecticut Department of Public Health, Hartford, United States; <sup>9</sup>California Cancer Reporting and Epidemiologic Surveillance Program, University of California, Davis, Davis, United States; <sup>10</sup>Cedars-Sinai Cancer and Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, United States; <sup>11</sup>The Iowa Cancer Registry, University of Iowa, Iowa City, United States; <sup>12</sup>Howard Hughes Medical Institute, University of Minnesota, Minneapolis, United States

\*For correspondence: [gabe.starrett@nih.gov](mailto:gabe.starrett@nih.gov)

†These authors contributed equally to this work

Competing interest: [See page 21](#)

Funding: [See page 21](#)

Preprinted: 11 November 2021  
 Received: 13 August 2022  
 Accepted: 22 March 2023  
 Published: 24 March 2023

Reviewing Editor: Nicholas Wallace, Kansas State University, United States

© This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0 public domain dedication](#).

**Abstract** A small percentage of bladder cancers in the general population have been found to harbor DNA viruses. In contrast, up to 25% of tumors of solid organ transplant recipients, who are at an increased risk of developing bladder cancer and have an overall poorer outcomes, harbor BK polyomavirus (BKPyV). To better understand the biology of the tumors and the mechanisms of carcinogenesis from potential oncoviruses, we performed whole genome and transcriptome sequencing on bladder cancer specimens from 43 transplant patients. Nearly half of the tumors from this patient population contained viral sequences. 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%). Immunohistochemistry revealed variable Large T antigen expression in BKPyV-positive tumors ranging from 100% positive staining of tumor tissue to less than 1%. In most cases of BKPyV-positive tumors, the viral genome appeared to be clonally integrated into the host chromosome consistent with microhomology-mediated end joining and coincided with focal amplifications of the tumor genome similar to other virus-mediated cancers. Significant changes in host gene expression consistent with the functions of BKPyV Large T antigen were also observed in these tumors. Lastly, we identified four mutation signatures in our cases, with those attributable to APOBEC3 and SBS5 being the most abundant. Mutation signatures associated with an antiviral drug, ganciclovir, and aristolochic acid, a nephrotoxic compound found in some herbal medicines, were also observed. The results suggest multiple pathways to carcinogenesis in solid organ transplant recipients with a large fraction being virus-associated.

## Editor's evaluation

This work provides a compelling case that viral origins of bladder cancer should be more carefully considered. Specifically, the clonogenic expression of viral oncogenes in these tumors combined with the lower than expected prevalence of major tumor suppressor (p53 and pRB) provides strong evidence for the authors' assertions. Ultimately, it will be important to follow up this work and I look forward to seeing those next steps.

---

## Introduction

At least 20% of all cancers are attributable to viral, bacterial, or parasitic infections (*de Martel et al., 2020*). 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 (*Zapatka et al., 2020; Cantalupo et al., 2018*). 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 (*Cantalupo et al., 2018*).

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. These populations have been previously shown to be at increased risk for developing papillomavirus-mediated cancers, and oncogenic viruses, such as Kaposi's sarcoma-associated herpesvirus and Merkel cell polyomavirus (MCPyV), were discovered in these patient populations (*Chang et al., 1994; Feng et al., 2008; D'Arcy et al., 2021*).

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 (*Graham, 2017*). The carcinogenic effects of these small circular double-stranded DNA viruses are primarily dependent on the expression of the E6 and E7 oncogenes which, among a wide range of other functions, inactivate the tumor suppressor proteins p53 and Rb, respectively (*Rosty et al., 2005; Crook et al., 1991; Mirabello et al., 2017; DeCaprio, 2014; Barbosa et al., 1990*).

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 (*Moens and Macdonald, 2019*). MCPyV has been identified as an etiological factor in a rare skin cancer, Merkel cell carcinoma (*Feng et al., 2008; Shuda et al., 2008; Kassem et al., 2008*). Another human 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 (*Gupta et al., 2018; Vajdic and van Leeuwen, 2009; Li et al., 2021; Papadimitriou et al., 2016*).

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 (*Robertson et al., 2017; Burns et al., 2013; Roberts et al., 2013*). These enzymes normally function as innate immune defenses against viruses by deaminating cytosines in single-stranded DNA, leading to hypermutation of the viral genome (*Poulain et al., 2020*). 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 (*Swanton et al., 2015*). APOBEC3A and APOBEC3B are upregulated in response to the expression of HPV E6 and E7, and APOBEC3A can restrict HPV replication (*Mori et al., 2017; Warren et al., 2017; Ahasan et al., 2015; Warren et al., 2015b; Vieira et al., 2014*). The large T antigens (LTags) of BKPyV and JC polyomavirus (JCPyV, a close relative of BKPyV) also upregulate APOBEC3B expression and activity (*Starrett et al., 2019; Peretti et al., 2018; Verhalen et al., 2016*).

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

Characteristic	Statistic	
	Median	IQR
Age in years at diagnosis	65	60, 71
Years from transplant to diagnosis	5.8	3, 7
	N	%
Sex		
Female	13	30
Male	30	70
Transplanted organ		
Kidney	24	56
Liver	4	9
Heart and/or lung	14	33
Pancreas	1	2
Race		
Non-Hispanic White	30	70
Asian/Pacific Islander	8	19
Hispanic	5	12
Summary stage		
In situ	12	28
Localized	19	46
Regional	7	14
Distant	5	12
Grade		
Low	20	47
High	22	51
Papillary urothelial neoplasm of low malignant potential	1	2

IQR: interquartile range.

11 remaining cases either had regional or distant invasion or metastasis. We successfully generated WGS data for 38 primary tumors, three metastases, and 10 normal (histologically non-malignant) tissues, with a median of 31 x coverage across the human genome (range: 14–55 x) (**Supplementary file 1a**). We generated total RNA sequencing data for 43 primary tumors, five metastases, and 14 normal tissues, with a median of 30 million reads per sample (range: 4–65.5 million).

### Detection of viruses in bladder cancers from transplant recipients

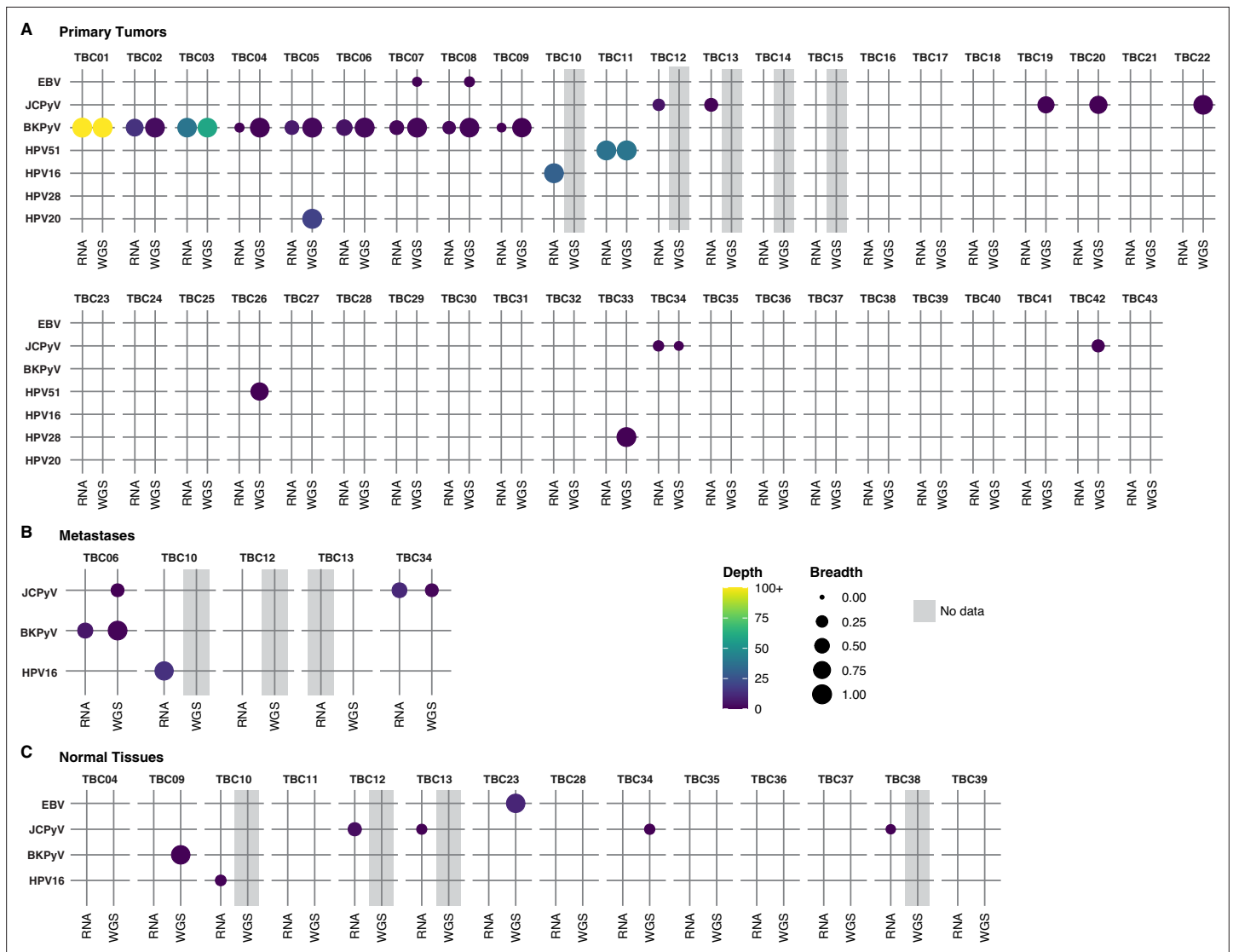
Analysis of WGS data for 38 primary tumors identified one or more virus species in 17 specimens (45%) (**Figure 1A** and **Supplementary file 1a**). RNA sequencing on tumor samples for which WGS data could not be obtained revealed three additional cases containing viral sequences (45% of samples overall). 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

To characterize the mutational, transcriptomic, and viral landscapes of 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 databases of 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.

## Results

### Bladder cancers from transplant recipients

The study population was comprised of 43 U.S. cases from patients who developed bladder cancer after receiving solid organ transplantation (**Table 1** and **Supplementary file 1a**). Seventy percent were male and 70% were non-Hispanic white. The median age at cancer diagnosis was 65 years (range: 27–82). The most commonly transplanted organ was the kidney (56%), followed by the heart and/or lung (33%) and liver (9%). Primary tumors were roughly an equal mixture of high- and low-grade carcinomas diagnosed with 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



**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. TBC numbers represent a single case and are consistent across primary, metastatic, and normal tissues. 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.

The online version of this article includes the following figure supplement(s) for figure 1:

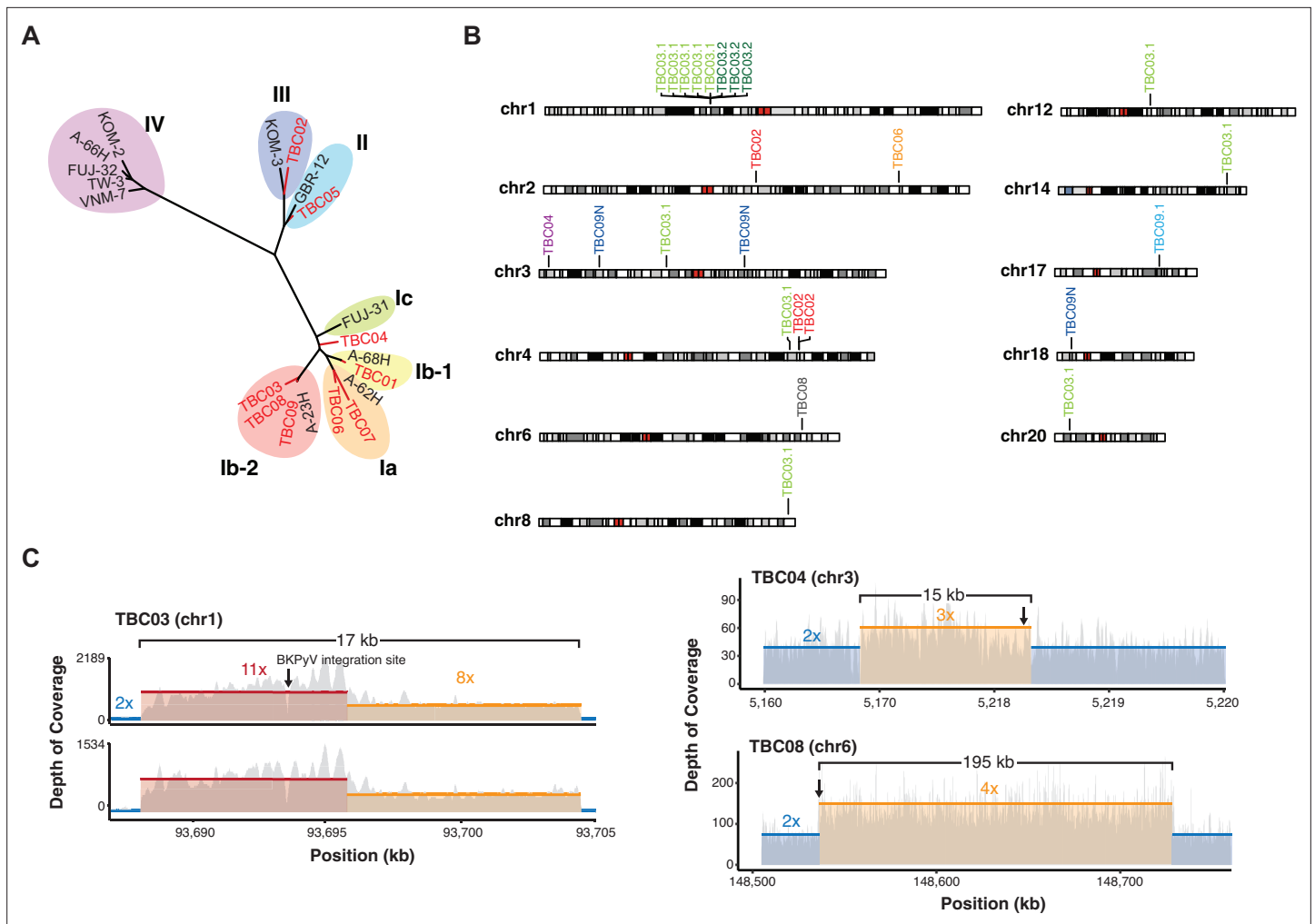
**Figure supplement 1.** BK polyomavirus (BKPyV) DNA/RNA coverage plots.

**Figure supplement 2.** JC polyomavirus (JCPyV) DNA/RNA coverage plots.

**Figure supplement 3.** All human papillomavirus (HPV) DNA/RNA coverage plots.

low-risk papillomavirus, HPV28, was observed in TBC33. One BKPyV-containing tumor (case TBC05) also harbored relatively abundant amounts of HPV20. Only two reads mapped to HPV20 in the RNA dataset for case TBC05. 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 (*Figure 1—figure supplement 1*).

WGS from TBC16, TBC17, TBC18, TBC19, TBC20, TBC21, TBC22, TBC23, TBC24, TBC27 had low numbers of reads mapping to the BKPyV genome that was 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.



**Figure 2.** Virus diversity and integration. **(A)** Phylogenetic tree of BK polyomavirus (BKPyV) large T antigen (LTag) sequences detected in tumors (red) and reference genotypes with representative strain names. **(B)** Sites of BKPyV integration into host chromosomes are indicated with case numbers. Two separate sections from separate formal-fixed paraffin-embedded (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 TBC03.1, TBC03.2, TBC04, and TBC08. BKPyV integration junctions are indicated by a black arrow. Colored numbers in the body of the graph indicate coverage depth.

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 (**Supplementary file 1c**). Epstein-Barr virus was most strongly detected in one normal lymph node (TBC23) and, at low levels, in tumors TBC07 and TBC08. Considering the stronger epidemiological evidence for BKPyV and bladder cancer and its abundance in these specimens, we focused the majority of our analysis on characterizing these tumors.

### Features of BKPyV-positive tumors

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. BKPyV-positive tumors were found in two heart, two lung, one heart and lung, and four kidney transplant recipients. 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 the case TBC03, two separate sections from separate blocks

**Table 2.** BK polyomavirus (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	CTCTGCTCATGAA	CATGAAGGT TAAGCATGCTA	5	ATGAA	chr4	145732354	C4orf51	ENSG00000237136	0	AluSq2	-474
TBC02	ACCATTAAATCCCAA	AGTGGAAATTAC	2	AC	chr4	145732375	C4orf51	ENSG00000237136	0	AluSq2	-495
TBC03.1	GCCTTCTTG TGGACTGGT	ATTTCATTCT ACTGGGTCAGGA	0	No overlap	chr1	93693546	BCAR3	ENSG00000137936	0	MIRb	377
TBC03.1	TCGTCTTCT TATTTTCAGAA	GGGTTCTCTG TTTATAAGGTC	2	TC	chr1	93693570	BCAR3	ENSG00000137936	0	MIRb	353
TBC03.1	AGAGCCTTG GTGGTGG	GGTGGCAA CAGTGCAG	5	GGTGG	chr1	93693890	BCAR3	ENSG00000137936	0	MIRb	33
TBC03.1	GATACTTTT AGACATC	AACATGACC TCAGGAAGGA	4	CATG	chr1	93694075	BCAR3	ENSG00000137936	0	MIRb	0
TBC03.1	CCTCAAAGC CACCACTCC	TTTCCATGA GCCCCAAA	5	CCAAA	chr1	93694843	BCAR3	ENSG00000137936	0	MER5A	-92
TBC03.1	CAATTTTTTTTTT	TTTTTTAT TGTAAAGGTG	7	TTTTTT	chr12	50449935	LARP4	ENSG00000161813	0	AluSc	0
TBC03.1	TGCAAGGTG CTTCATGTAT	AGGGGGCTTA AAGGATGCA	4	TGCA	chr14	95764390		ENSG00000257275	-6735	MIRb	0
TBC03.1	TAGCCAAA AAAAAAGG	AAAAAAA GGCCACAG	11	AAAAAAAAGG	chr20	8525269	PLCB1	ENSG00000182621	0	MamSINE1	154
TBC03.1	CAATTTGGA AAACAAT	ATGCAAGG CAGTGACA	2	AT	chr3	73059264	PPP4R2	ENSG00000163605	0	MER103C	69
TBC03.1	TAAAAAGTGTC	AAAGTGCAA TAGAGAAAA	8	AAGTGTC	chr4	142307350	INPP4B	ENSG00000109452	0	L2a	0
TBC03.1	TCACAAAT TT-TACTCTCT	ACACTTTTAC ACTCTCTA	8	ACTCTCT	chr8	140923993	PTK2	ENSG00000169398	0	L2a	0
TBC03.2	GTTGAGTT GGAGCA	CATCTAAATA TCTCTCAAAT	2	CA	chr1	93693160	BCAR3	ENSG00000137936	0	MER5A1	-10
TBC03.2	ACCAGTCCA CAAGAAAAGG	CCAGTAGAA ATGAAAAT	0	No overlap	chr1	93693546	BCAR3	ENSG00000137936	0	MIRb	377
TBC03.2	TCGTCTTCT TATTTTCAG	GTTCTCTGT TTATAAGGTC	2	TC	chr1	93693570	BCAR3	ENSG00000137936	0	MIRb	353
TBC04	GAGTGAGT TCATAG	CAACACTGTG GTGAG-TGAGTT	4	GAGT	chr3	5202593	EDEM1	ENSG00000134109	0	L2b	-466
TBC06	CAGACATT -AGGA	TGAGGACC TAACCTGT	4	AGGA	chr2	201676427	MPP4	ENSG00000082126	0	MIR1_Amn	0
TBC08	TCCACTTT CAGTACTT	TGCAAAA AATCAAAT	1	T	chr6	148535326	SASH1	ENSG00000111961	0	AluSq	995
TBC09.1	GGGGCGG TAACTAGAAG	ACTAGAAG CTTGTCTG	8	ACTAGAAG	chr17	61340185	BCAS3	ENSG00000141376	0	L2-3_Crp	0
TBC09N	GAGAAAAT AGGACTCGG	AAGATTCCG CTGAGAAA	7	GAGAAA	chr18	8169205	FTPRM	ENSG00000173482	0	MER127	-648

Table 2 continued on next page

Table 2 continued

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
TBC09N	TCCATCC TCCTCTAC	CTCCTCT ACATTGT	9	CTCCTCTAC	chr3	34028749	LINC01811	ENSG00000226320	130585	L2b	0
TBC09N	ATGTAAT ATAAACT	CATGATT TTAACCCAG	0	No overlap	chr3	117678477		ENSG00000239268	0	L2c	0

MH: microhomology; RE: Repeat element.



of the primary tumor were sequenced. In 1 of the 2 sections, 11 integration junctions were identified across seven chromosomes. 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 2 bp. In this model, which has previously been proposed for both HPV- and MCPyV-associated tumors ([Starrett et al., 2020](#); [Starrett et al., 2017](#); [Akagi et al., 2014](#)), 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. Consistent with this model, 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 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.

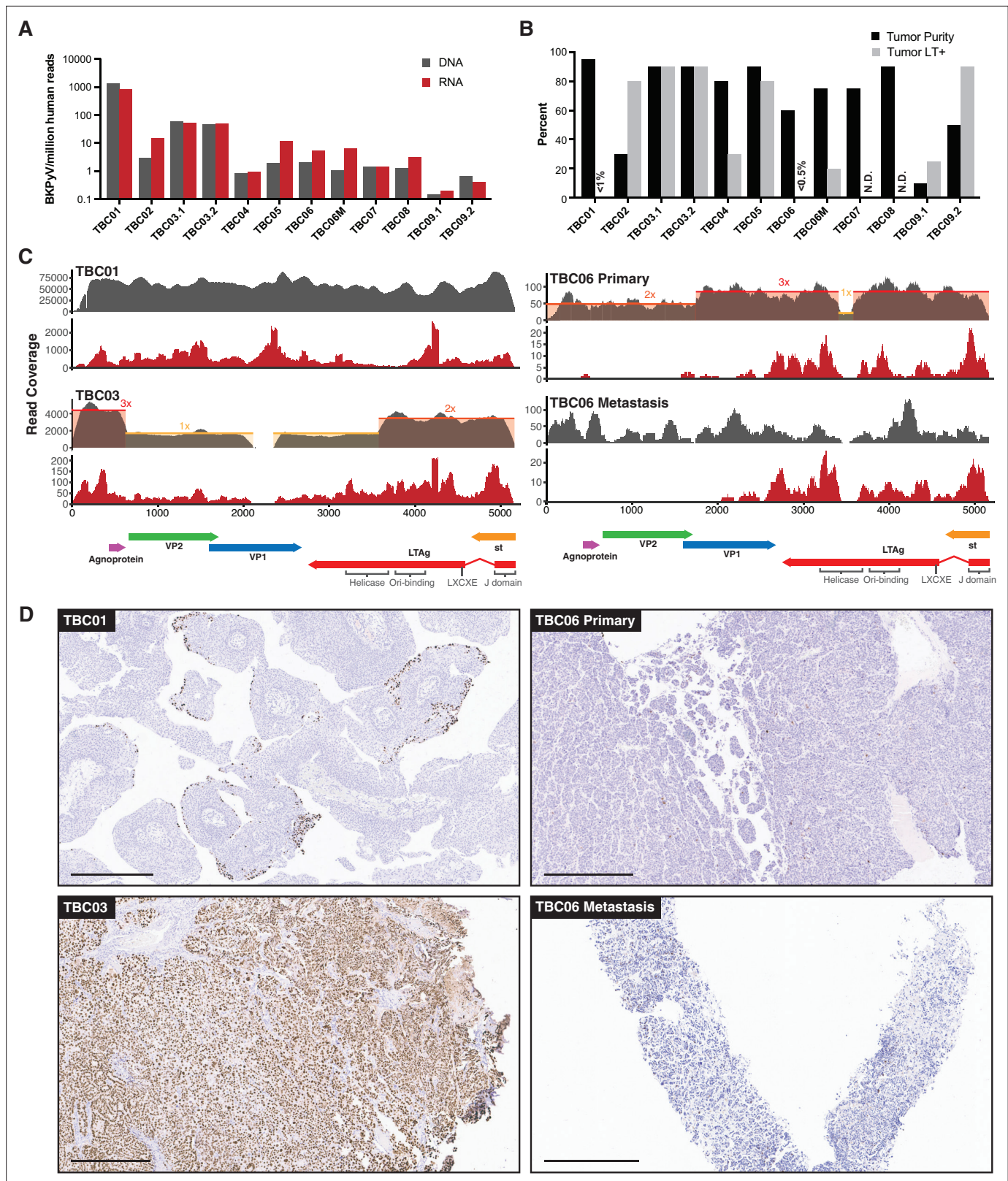
BKPyV RNA and DNA abundance by sequencing generally did not correspond to specimen tumor purity or the percentage of LTag + cells ([Figure 3A and B](#)). Gene-level analysis of the RNA sequencing data revealed that 7 of the 9 polyomavirus-positive tumors predominantly expressed the T antigens, with little to no expression of the late genes VP1 and VP2 (encoding the major and minor capsid proteins, respectively) ([Figure 3A and C](#), [Figure 3—figure supplement 1](#)). The LTag open reading frames (ORFs) in these cases were truncated before the helicase domain through deletions, frame-shifts, or point mutations. The exception was the BKPyV-positive PUNLMP case TBC01, which showed a 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. However, in this study, TBC01 was the only polyomavirus-positive tumor showing evidence of regulatory region rearrangements ([Figure 3—figure supplement 2](#)).

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 [Figure 3—figure supplement 3](#)). 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% of 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 ([Figure 3—figure supplement 3](#)). 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 as determined by cell morphology). LTag-positive cells in the TBC01 sample were almost entirely localized to the luminal margin of the tumor ([Figure 3D](#)).

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 4A](#), [Supplementary file 1e](#)). Clustering all primary and metastatic tumors by genes with a greater than threefold 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 4C](#)). 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



**Figure 3.** BK polyomavirus (BKPv) DNA, RNA, and large T antigen (LTag) detection in tumors. (A) Barplots showing the abundance of BKPv DNA and RNA reads standardized to human reads (B) Barplots of histologically estimated percent tumor purity and Immunohistochemistry (IHC)-positivity for polyomavirus LTag expression. N.D. indicates no IHC image data were generated. (C) Representative coverage plots for BKPv DNA (gray) and RNA

Figure 3 continued on next page

Figure 3 continued

(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 with scale bars representing 500 microns.

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** BK polyomavirus (BKPyV) gene expression.

**Figure supplement 2.** Diagrams of the assembled BK polyomavirus (BKPyV) NCCR structures and rearrangements in tumors.

**Figure supplement 3.** T antigen immunohistochemistry (IHC) in BK polyomavirus (BKPyV)-positive tumors.

proliferation in bladder cancer, such as *FGFR3*, had significantly lower expression in BKPyV-positive tumors relative to virus-negative tumors. Notably, tumors harboring BKPyV had significantly higher average *APOBEC3B* expression compared to both normal tissues and tumors not containing any virus (**Figure 4B**). 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 the highest average *APOBEC3B* expression observed in tumors with both BKPyV and two copies of rs1014971 (**Figure 4—figure supplement 1**).

In TBC03, the observed BKPyV integration into *BCAR3* results in increased expression of the host gene. Further evaluation of RNA reads covering this region revealed 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 (**Figure 4—figure supplement 2**). 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.

Aside from integration-related copy number variants (CNVs), large-scale CNVs overall differed between BKPyV-positive and virus-negative tumors (**Figure 5; Supplementary file 1g**). BKPyV-positive tumors showed moderate enrichments for gains of chromosome segments 1q, 2 p, 3 p, 7q, 20q, and 22q, while losses of chromosome 2q, 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 (**Starrett et al., 2020**).

## Features of other virus-positive tumors

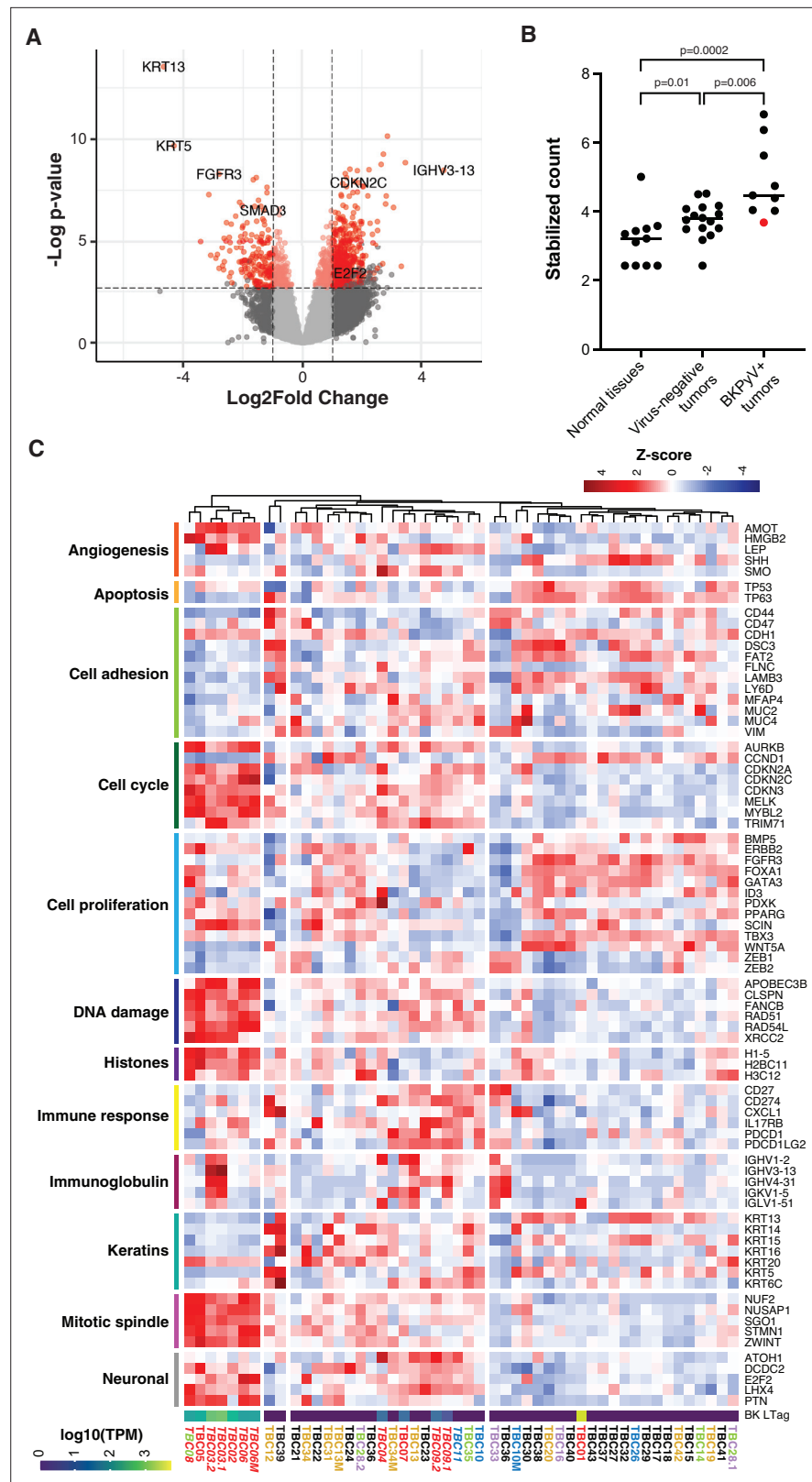
In the cases that were positive for JCPyV, DNA, and RNA coverage depth was much lower than observed for BKPyV-positive tumors, and in several DNA-positive cases, JCPyV transcription was not detected (**Figure 1**). JCPyV reads were detected in three samples from case TBC12 including the primary tumor, tumor-positive lymph node, and adjacent normal bladder wall (**Figure 1—figure supplement 2**). IHC detected sparse LTag staining in JCPyV-positive case TBC13, but not in any tissue samples for TBC12.

For 2 of 3 cases harboring HPV types known to cause cervical cancer (HPV16 and HPV51), transcripts encoding the E6 and E7 oncogenes were detected (**Figure 1—figure supplement 3**). In one HPV16+ case (TBC10), viral oncogene RNA expression was detected in both the primary and metastatic specimens. A possible HPV51 integration event in TBC11 appears to have involved simple repeat sequences and retroelements (**Figure 1—figure supplement 3**). Lastly, one case harbored DNA sequences aligning to HPV20 and a single case harbored DNA aligning to HPV28; however, no RNA reads were detected for these cutaneous HPV types (**Figure 1—figure supplement 3**).

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 (**Tisza et al., 2020**). All observed TTV ORF1 sequences belonged to the *Alphatorquevirus* genus and had 51–100% amino acid identity to previously reported TTV strains (**Supplementary file 1d**).

## Mutation signature analysis

The overall tumor mutation burden, as measured by non-synonymous mutations per million bases, did not show a clear correlation with the presence of viral sequences (**Figure 6A**). We analyzed likely somatic point mutations from all tumors and deconvoluted mutation signatures (**Figure 6B and C, Figure 6—figure supplement 1**). As expected for bladder cancer, we commonly observed single-base



**Figure 4.** Differential gene expression in BK polyomavirus (BKPyV)-positive tumors. **(A)** Volcano plot of differential gene expression between BKPyV-positive and virus-negative BK tumors. Significantly differentially expressed genes ( $q\text{-value} < 0.05$ , DESeq2) with a fold change greater than two are in red, and genes with a fold change less than two are in pink. Non-significant genes are in gray. **(B)** Variance stabilized counts for APOBEC3B expression from *Figure 4 continued on next page*

Figure 4 continued

DESeq2 grouped by normal tissues, virus-negative tumors, and BKPyV-positive tumors showing significantly increased expression in BKPyV-positive tumors (Mann-Whitney U test). TBC01 is indicated by a red dot. (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 names are colored by likely etiology: BKPyV-positive, red; JC polyomavirus (JCPyV)-positive, goldenrod; HR-HPV-positive, blue; torque teno virus (TTV)-positive, green; aristolochic acid, purple; undetermined, black; multiple colors reflect multiple detected viruses or etiologies. Tumors with evidence of integration are in italics. BKPyV LTag expression is shown as  $\log_{10}(\text{transcripts per million [TPM]})$ .

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** APOBEC3B germline variant and expression by BK polyomavirus (BKPyV) status.

**Figure supplement 2.** Host transcripts from the BK polyomavirus (BKPyV) integration site at BCAR3.

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).

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 6A**). 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, which was recently identified in hematopoietic stem cell transplant recipients (**Figure 6B; de Kanter et al., 2021**).

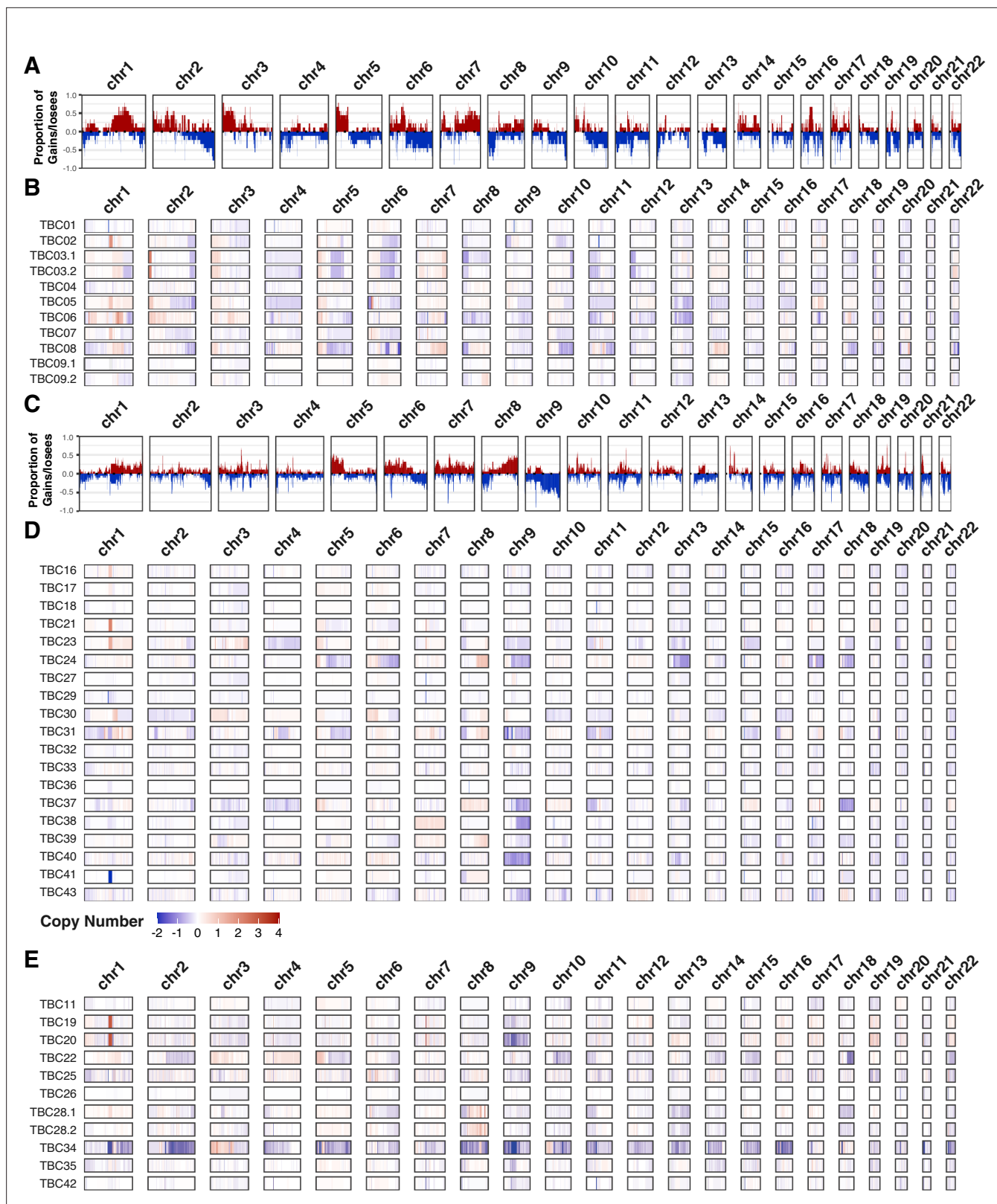
## Recurrent somatic mutations

First, 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 6D**). 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 6D**). 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.

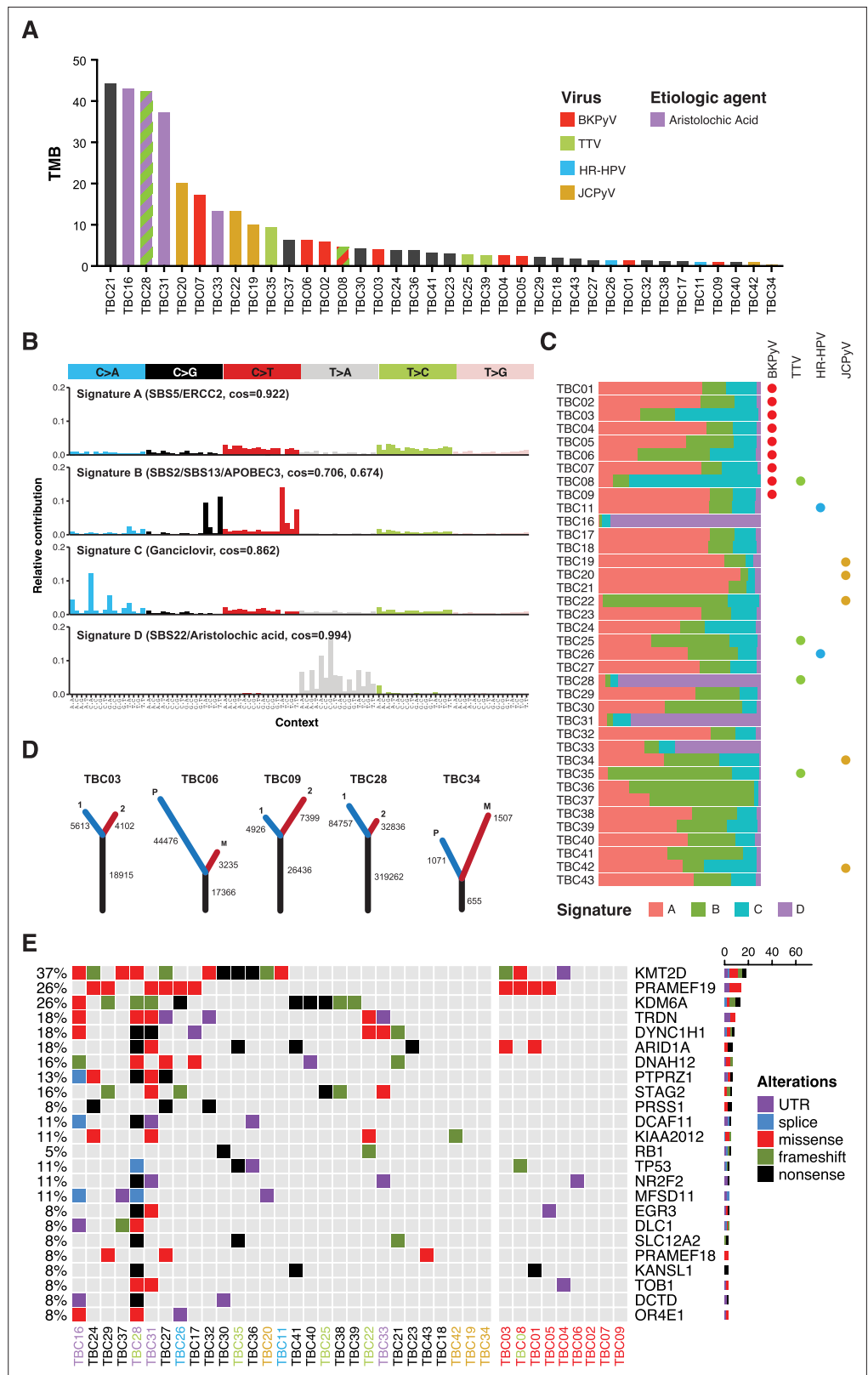
Numerous cellular genes were found to recurrently harbor nonsynonymous, nonsense, and frameshift mutations (**Figure 6E**). The spectrum of frequently mutated genes is similar to those reported in various types of urothelial carcinoma (e.g. mutations in *KMT2D*, *KDM6A*, and *ARID1A*) (**Supplementary file 1f; Robertson et al., 2017; Nassar et al., 2019; Su et al., 2021**). No nonsynonymous mutations were identified in *FGFR3* or *PIK3CA*, even though these genes are commonly mutated in non-muscle invasive bladder cancer (NMIBC) (**Cancer et al., 2014**). Mutations in *TP53*, which are common in muscle-invasive bladder cancer (**Robertson et al., 2017**), were detected in four tumors (**Figure 6E**). None of the HPV-positive tumors with WGS harbored mutations in *TP53* or *RB1*. Similarly, none of the polyomavirus-positive tumors harbored mutations in *RB1* and only TBC08 had a frameshift mutation in *TP53*.

## 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 10 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 (**Cantalupo et al., 2018; Llewellyn et al., 2018**). The results suggest that human polyomaviruses



**Figure 5.** Copy number variants. Frequency plots for large copy number variants in BK polyomavirus (BKPyV)-positive tumors (panel **A**) and virus-negative tumors (panel **C**). Frequency of gains/amplifications is shown in red; losses/deletions are shown in blue. Sample level copy number variant spectra for BKPyV-positive tumors (panel **B**), virus-negative tumors (panel **D**), and all other tumors (panel **E**). Complete deletions are in dark blue and high copy amplifications are in red.



**Figure 6.** 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, BK polyomavirus (BKPyV); green, TTV; blue, HR-HPV; goldenrod, JC polyomavirus (JCPyV)) or etiologic agent (aristolochic acid, purple; black, undetermined). Multiple colors reflect multiple detected viruses or etiologies. **(B)** Figure 6 continued on next page

Figure 6 continued

Barplots of the contribution of 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 one and two 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; undetermined, black. The percent of modified tumors is shown on the left and the count of the variants in each gene is represented by the barplot on the right.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Mutations signature deconvolution.

and papillomaviruses can play a carcinogenic role in the development of bladder cancer, particularly among transplant recipients.

BKPyV infection in organotypic urothelial cell culture has been shown to promote cellular proliferation (Schneidewind et al., 2020). This is most likely through the transforming effects of viral T antigens, as is supported here and in previous studies by the loss of late region transcripts and enrichment of early region transcription in primary tumors, and maintenance of these transcripts in metastatic lesions (Müller et al., 2018). Interestingly, we observed frequent clonal loss of the p53-inactivating helicase domain of BKPyV LTag due to deletions and point mutations in the integrated virus. While such deletions in LTag are commonly observed in MCPyV-positive Merkel cell carcinoma (Shuda et al., 2008), 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 LTag to be preserved in tumor cells. We speculate that the loss of the BKPyV helicase domain is driven by negative selection against deleterious effects of LTag 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). The absence of the p53-binding domain may be compensated for in some BKPyV-positive tumors through the significantly increased expression of the ubiquitin ligase TRIM71 that we observed. TRIM71 has been shown to bind and poly-ubiquitinate p53 for proteasomal degradation and prevent apoptosis during stem cell differentiation (Nguyen et al., 2017).

We also observed amplification of the host genome surrounding BKPyV integration sites, consistent with circular DNA intermediates and/or MMEJ break-induced replication. Similar findings have been reported for HPV and MCPyV-associated tumors (Starrett et al., 2020; Czech-Sioli et al., 2020). These amplification events result in a variable number of tandem head-to-tail copies of the virus and host genome that are thought to create super-enhancers affecting viral and host gene expression (Warburton et al., 2018; Dooley et al., 2016). In cervical cancer, frequently only one integration event is transcriptionally active; however, in tumors carrying integrated BKPyV sequences, the abundance of viral DNA and RNA are positively correlated, suggesting that each integrated copy produces viral transcripts. 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 (Jeannot et al., 2018). Elevated expression of BCAR3 has been shown to increase the proliferation, motility, and invasiveness of estrogen receptor-positive breast cancer cells after treatment with anti-estrogens (Wallez et al., 2014; Wilson et al., 2013).

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 high-grade and invasive behavior of BKPyV-positive tumors observed in this study and others (Alexiev et al., 2013; Sirohi et al., 2018; Kenan et al., 2015; Kenan et al., 2017; Nickeleit et al., 2018).

Many of the observed gene expression changes are consistent with known effects of BKPyV infection and the specific activities of LTag, which binds Rb-family proteins and alters the active pool of E2F transcription factors in the cell (Caller et al., 2019; Harris et al., 1996). Recent studies have shown that APOBEC3B expression is repressed by the DREAM complex (which is composed of Rb-family proteins



and E2F transcription factors **Starrett et al., 2019**) and, accordingly, we found that APOBEC3B is more highly expressed in BKPyV-positive tumors compared to normal tissues and tumors without tumor virus sequences, likely due to LTag activity. However, despite this increased expression, the mutation signature commonly attributed to APOBEC3B did not appear enriched in BKPyV-positive tumors. It is possible that tumors expressing BKPyV LTag and increased APOBEC3B manifested greater intratumor mutational heterogeneity, but we were unable to detect possible low-frequency APOBEC3-mediated variants from FFPE tissue without deeper and more accurate sequencing. Additionally, consistent with the disruption of the DREAM complex in these tumors, we observed higher expression of *MYBL2*, a key component of the MMB complex, and one of its targets, *FOXM1*, which regulates numerous genes required for G2/M progression. We also observed increased expression of *FOXM1* downstream targets associated with the centromere and kinetochore, which have been shown to promote improper chromosome segregation and tumorigenesis (**Fischer and Müller, 2017; Sadasivam et al., 2012; Schade et al., 2019**).

BKPyV-positive tumors in our study had significantly higher expression of a number of genes that promote homologous recombination (e.g. *RAD51*, *RAD54L*, *BRCA1*, and *BRCA2*) and protect against replication fork stalling and collapse (e.g. *RAD51*, *XRCC2*, and *FANCB*) relative to virus-negative tumors (**Tye et al., 2021**). Claspin (*CLSPN*) and *TIMELESS*, which interact with replicative polymerases and helicases, are also highly expressed in BKPyV-positive tumors, further promoting replication fork progression and genome stability (**Bianco et al., 2019**). This expression pattern might promote cell survival in the face of genomic damage caused by viral genome integration, oncogene expression, and APOBEC3B upregulation.

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 (**Zapatka et al., 2020; Cantalupo et al., 2018; Guma et al., 2016**). In the current study, we identified four tumors with carcinogenic *Alphapapillomavirus* sequences (HPV16 or HPV51). *Alphapapillomaviruses* are believed to cause cancer through the sustained expression of their E6 and E7 oncoproteins, which is frequently associated with the integration of the papillomavirus genome into the tumor genome.

One case in the panel carried sequences of HPV20, a *Betapapillomavirus* that can cause cutaneous squamous cell carcinoma in animal model systems (**Michel et al., 2006**). The possible involvement of *Betapapillomaviruses* in skin cancer in the general population remains controversial (**Viarisio et al., 2018**). 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* (**Dell'Oste et al., 2009**). 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 (**Michel et al., 2006; Viarisio et al., 2018; Viarisio et al., 2016**). In the context of bladder cancer, it is possible that cutaneous papillomaviruses likewise enable the accumulation of carcinogenic DNA damage. Additionally, the identification of HPV28, an *Alphapapillomavirus* that is not generally associated with cervical cancer, suggests more abundant papillomavirus infections of the bladder than previously assumed, with unknown implications for carcinogenesis.

An 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, in combination with immune suppression, 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 (**Solis et al., 2018**), but has not been documented for heart, lung, or liver transplant recipients, who are also included in the current study. Additionally, this study and one prior study (**Querido et al., 2020**) identified JCPyV in bladder tumors. Based on the high degree of similarity between JCPyV and BKPyV, it seems reasonable to expect that the two species would behave similarly. However, the low abundance of JCPyV RNA and DNA in these specimens and the absence of integration, together with the ubiquity of latent JCPyV infections in the urinary tract, raises the possibility that these observations reflect incidental detection events.

The data from this study and others suggest that in the context of strong immune suppression BKPyV can cause bladder cancer through clonal integration but is rarely detected in tumors of the general population. While most adults are seropositive for BKPyV, with at least 10% having detectable

BKPyV in the urine, BKPyV is only observed in upwards of 4% of NMIBC cases and less than 0.25% in muscle-invasive bladder cancers in the general population ([Cantalupo et al., 2018](#); [Llewellyn et al., 2018](#)). This implies that, 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 Merkel cell carcinomas. Immunologic recognition of these tumors may also be impacted by the increased expression of APOBEC3B, which can generate immunogenic neoantigens ([Serebrenik et al., 2019](#); [Chen et al., 2019](#)). 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 ([Meier et al., 2021](#); [Cuenca et al., 2020](#); [Fu et al., 2018](#)). The theoretical immunological costs of viral gene expression for a nascent tumor cell raise the possibility of 'hit-and-run' carcinogenesis. The hit-and-run hypothesis invokes the idea that a virus may play a causal role in the 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, the expression of viral oncogenes may promote genome instability through the expression of the mutagenic APOBEC3 enzymes or other mechanisms that further push the cell towards transformation, as has been suggested by a recent study of BKPyV infection in differentiated urothelium ([Baker et al., 2022](#)).

The heterogeneous expression of LTag observed in this study could represent transcriptional silencing or loss of BKPyV DNA from one part of the tumor, supporting the idea that tumors can lose the need for LTag expression. Alternatively, our observations could be accounted for by a multi-stage integration and carcinogenesis process proposed by other recent studies on BKPyV-positive urinary tumors from kidney transplant recipients ([Jin et al., 2021](#); [Wang et al., 2020](#)). However, our sequencing experiments support a dominant clonally integrated form likely established early during tumor development 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 tumor likely represents a passenger infection of an existing tumor ([Dalianis and Hirsch, 2013](#)). Future studies should investigate the hypothesis that passenger infections might play an oncomodulatory role in tumor development. This also suggests that archetypal BKPyV, rather than the more pathogenic rearranged strains found in cases of nephropathy, is more likely to integrate and be preserved into nascent tumor cells. In support of the idea that integration may be a more common aspect of BKPyV infection than previously assumed, we identified a clonal BKPyV integrant in the normal bladder specimen from case TBC09 in both the RNA and WGS sequencing that was distinct from the BKPyV integrant observed in the tumor sample. The normal tissue integrant had multiple copies of small T antigen and a large deletion in the regulatory regions ([Figure 1—figure supplement 1](#)). Only a few reads from RNA sequencing mapped to the small T antigen region, and histology of the section indicates no tumor cells or LTag staining, suggesting that the virus did not integrate into the right genomic location or maintain the needed components to drive carcinogenesis.

It remains to be seen whether TTVs contribute to disease in the context of immune suppression. A general model is that these ubiquitous viruses 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 in the diversity of TTVs observed in hosts ([De Vlamincx et al., 2013](#)). Indeed, the detection of TTVs can serve as an indicator of the degree of overall immune suppression in transplant recipients ([Blatter et al., 2018](#)). 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 ([Poulain et al., 2020](#); [Verhalen et al., 2016](#); [Warren et al., 2015a](#)).

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 sequencing ([Robbe et al., 2018](#)). 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 limitation of this work, but our analytical approach accounted for this 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 (Zhang *et al.*, 2014). Our variant calling approach was also validated by the observation that we detected four deconvoluted mutation signatures that match those expected from prior surveys of bladder cancer. However, the low overall coverage of our WGS remains a limitation of this study.

We identified four 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, which sometimes contaminates certain types of herbal medicines and grains (Poon *et al.*, 2015). 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 four cases with dominant aristolochic acid signatures were in the top seven for total mutation burden (Figure 6). None of the three tumors had detectable oncogenic viral sequences, but one had detectable TTV. We also identified likely ganciclovir-mediated mutations (de Kanter *et al.*, 2021) in most patients indicating that this common treatment to prevent reactivation of cytomegalovirus in solid organ transplant recipients may promote mutagenesis in the urinary bladder. Unfortunately, ganciclovir treatment history was unavailable to confirm that this is the origin of this mutation signature in these cases. Ever-decreasing sequencing costs will facilitate additional studies of this type and shed light on rare and understudied tumor types, as well as analyses of lower-grade and pre-cancerous lesions.

## 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 paraffin-embedded (FFPE) blocks for each specimen 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 and one additional section was used for immunohistochemistry. Hematoxylin and eosin-stained sections were reviewed by a trained pathologist and tumor purity was determined by cellular morphology.

### Nucleic acid isolation

Samples were simultaneously deparaffinized and digested using 400  $\mu$ L molecular-grade mineral oil (Millipore-Sigma) and 255  $\mu$ L Buffer ATL (Qiagen) supplemented with 45  $\mu$ L of proteinase K (Qiagen). Samples were incubated overnight at 65 °C in a shaking heat block. Samples were spun at 16,000  $\times$  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 2 hr long digests by the addition of 25  $\mu$ L of fresh proteinase K buffer. Lysates were stored at 4 °C until RNA or DNA isolation and processed within one month.

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

(Thermo Fisher) and spectrophotometry (DeNovix). DNA was stored at  $-20^{\circ}\text{C}$  until used for library preparation. Only samples with greater than 50 ng of DNA were processed for library prep.

For RNA isolation, 150  $\mu\text{L}$  of the remaining clarified lysate was moved to a new tube. 250  $\mu\text{L}$  of buffer PKD (Qiagen) was added and vortexed to mix. The remainder of the RNA extraction process was carried out using an 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  $\mu\text{m}$  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, including SV40, BKPyV, JCPyV, WU, KI, 6, 7, 10, and 11 (*Toptan et al., 2016*). Slides were baked in a laboratory oven at  $60^{\circ}\text{C}$  for 1 hr prior to immunostaining on Ventana Discovery Ultra automated IHC stainer upon following conditions: CC2 (pH9) antigen retrieval for 64 min at  $96^{\circ}\text{C}$ , antibody at concentration 0.5  $\mu\text{g}/\text{ml}$  in Agilent antibody diluent (cat. S3022) for 32 min at  $36^{\circ}\text{C}$ , Anti-Mouse HQ-Anti HQ HRP detection system for 12 min with DAB for 4 min and Hematoxylin II counterstain for 8 min. 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. 351 S) 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 s. 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 hr at  $20^{\circ}\text{C}$  then treated with 4  $\mu\text{L}$  USER enzyme (NEB) for 15 min at  $37^{\circ}\text{C}$  to digest uracil-containing fragments. Ligation reactions were cleaned up using 0.8 x 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 1 x 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's instructions for FFPE tissues. Final RNA and DNA libraries were assessed for size and quantity by Agilent TapeStation. Only samples that yielded libraries greater than 5 nM were sequenced.

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 *Supplementary file 1b*.

## Sequence alignments

Reads were trimmed using Trim Galore 0.6.0 with default settings. RNA reads was initially aligned using STAR aligner 2.5.3ab (*Dobin et al., 2013*) against a fusion reference human genome containing hg38, all human viruses represented in RefSeq as of December 2018 (*Supplementary file 1c*), and all papillomavirus genomes from PaVE <https://pave.niaid.nih.gov> (*Van Doorslaer et al., 2017*). 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 (*Langmead and Salzberg, 2012*). 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 (*Li et al., 2015*). All trimmed RNA reads were assembled using

RNASPAdes (*Bushmanova et al., 2019*). Assembled contigs were annotated using BLASTn and BLASTx against the NCBI nt database (October 2021) for closely related species, and Cenote-Taker2 version 2.1.2 (<https://github.com/mtisza1/Cenote-Taker2>; *Tisza, 2021*) was used to identify more divergent species in contigs  $\geq 1000$  bp (*Tisza et al., 2021*). Depth and breadth of coverage of viral species were normalized by the total number of human reads and length of the viral genome. Only species with  $\geq 10\%$  genome coverage and a normalized depth  $\geq 10$  for a viral genome 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 (*Martí-Carreras et al., 2020*).

Bam alignments were input into Oncovirus Tools ([https://github.com/gstarrett/oncovirus\\_tools](https://github.com/gstarrett/oncovirus_tools)) to call integration sites (*Starrett et al., 2020*; *Starrett, 2020*). It starts by extracting discordant read pairs (where one read aligns to a sequence of interest, i.e., virus, and the mated read aligns to the human genome) and any remaining reads aligned to the human genome that contain at least one 25 bp k-mer from the input sequences of interest as determined by a Bloom filter. It uses the human genomic coordinates from the above reads to identify putative integration regions by merging their stranded mapping positions to find overlaps, counting the number of reads per stranded region. Oncovirus Tools then assembles the extracted reads, together with all unaligned reads, using Spades (*Bankevich et al., 2012*). The resulting assembly graphs are annotated with the human and viral genomes using BLASTn and the annotated assembly graphs are plotted using the R package ggraph. The output is then screened for contigs containing both human and viral hits with BLASTn e-values below  $1e-10$ . Based on these hits, integration junctions are called and overlaps in host-virus hit on the contigs are then screened for microhomology. All putative integration sites from Oncovirus Tools were manually validated by returning to the original alignment file.

## Transcriptome clustering and differential gene expression analysis

Counts from STAR were input into R and normalized using the DESeq2 vst function (*Love et al., 2014*). The DESeq2 model was built using the following factor: tissue type (normal, primary, metastasis), grade, stage, and virus status to evaluate their effects on gene expression. Since the RNA seq libraries were prepared in different batches on different days and in different sequencing runs, 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>) (*Kuleshov et al., 2016*; *Chen et al., 2013*).

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

Point mutations were called using Mutect2, VarScan2, and lofreq with default parameters (*Koboldt et al., 2012*; *Wilm et al., 2012*). Consensus calls between these variant callers were performed using SomaticSeq (3.3.0) (*Fang et al., 2015*). Likely germline variants were annotated and removed using SnpSift and dbSNP v152. Likely somatic point mutations were further filtered by the following criteria: SomaticSeq PASS filter,  $\geq 10\%$  allele frequency,  $\geq 4$  reads supporting the variant allele, and  $\geq 8$  reads of total coverage of that position. Common mutations in cancer were annotated using SnpSift and COSMIC. Somatic mutations enrichment by gene was determined using the R package dNdScv. 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 were 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.

## Data visualization

All graphs were made in the R statistical environment (4.0.3) using the package ggplot2 or using GraphPad Prism.

## Acknowledgements

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

---

## Additional information

### Competing interests

Yelena Golubeva, Petra Lenz: is affiliated with Leidos Biomedical Research Inc. The author has no financial interests to declare. Reuben S Harris: is a co-founder, shareholder, and consultant of ApoGen Biotechnologies Inc. The other authors declare that no competing interests exist.

### Funding

Funder	Grant reference number	Author
National Institutes of Health	Intramural Research Program	Gabriel J Starrett Eric A Engels Christopher B Buck

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Gabriel J Starrett, Conceptualization, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Kelly Yu, Data curation, Validation, Project administration, Writing – review and editing; Yelena Golubeva, Mary L Piaskowski, Investigation, Writing – review and editing; Petra Lenz, Formal analysis, Validation, Writing – review and editing; David Petersen, Validation, Investigation, Writing – review and editing; Michael Dean, Ajay Israni, Brenda Y Hernandez, Thomas C Tucker, Iona Cheng, Lou Gonsalves, Cyllene R Morris, Shehnaz K Hussain, Charles F Lynch, Reuben S Harris, Ludmila Prokunina-Olsson, Paul S Meltzer, Resources, Writing – review and editing; Christopher B Buck, Conceptualization, Resources, Supervision, Funding acquisition, Writing – original draft, Project administration, Writing – review and editing; Eric A Engels, Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Writing – original draft, Project administration, Writing – review and editing

### Author ORCIDs

Gabriel J Starrett  <http://orcid.org/0000-0001-5871-5306>  
Mary L Piaskowski  <http://orcid.org/0000-0001-8453-6416>  
Reuben S Harris  <http://orcid.org/0000-0002-9034-9112>  
Ludmila Prokunina-Olsson  <http://orcid.org/0000-0002-9622-2091>  
Christopher B Buck  <http://orcid.org/0000-0003-3165-8094>

### Ethics

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.

**Decision letter and Author response**Decision letter <https://doi.org/10.7554/eLife.82690.sa1>Author response <https://doi.org/10.7554/eLife.82690.sa2>**Additional files****Supplementary files**

- Supplementary file 1. Supporting information. (a) Checkerboard table of samples used in this study. (b) Sequencing metrics. (c) Reference sequences used in this study. (d) Tumor torque teno virus similarities. (e) BKPvV-positive tumor vs virus-free tumor significantly differentially expressed genes. (f) Non-synonymous point mutations. (g) Copy number variants.
- MDAR checklist

**Data availability**

All refseqs for human papillomaviruses were downloaded from PaVE and refseqs for human polyomaviruses were downloaded from NCBI as of November 2018. All sequencing data generated in this study are available from dbGaP under accession phs003012.v1.p1. Viral contigs from this study are deposited in GenBank under accessions OQ469311 and OQ469312. All other contigs and larger IHC images are deposited in figshare ([https://figshare.com/projects/Common\\_Mechanisms\\_of\\_Virus-Mediated\\_Oncogenesis\\_in\\_Bladder\\_Cancers\\_Arising\\_in\\_Solid\\_Organ\\_Transplant\\_Recipients/132833](https://figshare.com/projects/Common_Mechanisms_of_Virus-Mediated_Oncogenesis_in_Bladder_Cancers_Arising_in_Solid_Organ_Transplant_Recipients/132833)). Code used in this manuscript are available from [https://github.com/gstarrett/oncovirus\\_tools](https://github.com/gstarrett/oncovirus_tools) (Starrett, 2020).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Starrett GJ	2022	Multi-omics of Bladder Cancers of Solid Organ Transplant Recipients	<a href="https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003012.v1.p1">https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003012.v1.p1</a>	NCBI dbGaP, phs003012.v1.p1
Starrett GJ	2023	pAb416 IHC images	<a href="https://doi.org/10.6084/m9.figshare.19199723.v1">https://doi.org/10.6084/m9.figshare.19199723.v1</a>	figshare, 10.6084/m9.figshare.19199723.v1
Starrett GJ	2023	Anelloviridae contigs	<a href="https://doi.org/10.6084/m9.figshare.22124240.v1">https://doi.org/10.6084/m9.figshare.22124240.v1</a>	figshare, 10.6084/m9.figshare.22124240.v1
Starrett GJ	2023	Anellovirus sp. isolate TBC35T_k141_20313 ORF2/4 (ORF2/4) gene, complete cds; and ORF1 (ORF1) gene, partial cds	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/OQ469311">https://www.ncbi.nlm.nih.gov/nucleotide/OQ469311</a>	NCBI Nucleotide, OQ469311
Starrett GJ	2023	Anellovirus sp. isolate TBC35T_k141_22964 ORF1 (ORF1) gene, partial cds; and ORF2 (ORF2) gene, complete cds	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/OQ469312">https://www.ncbi.nlm.nih.gov/nucleotide/OQ469312</a>	NCBI Nucleotide, OQ469312

**References**

- Ahasan MM, Wakae K, Wang Z, Kitamura K, Liu G, Koura M, Imayasu M, Sakamoto N, Hanaoka K, Nakamura M, Kyo S, Kondo S, Fujiwara H, Yoshizaki T, Mori S, Kukimoto I, Muramatsu M. 2015. Apobec3A and 3C decrease human papillomavirus 16 pseudovirion infectivity. *Biochemical and Biophysical Research Communications* **457**:295–299. DOI: <https://doi.org/10.1016/j.bbrc.2014.12.103>, PMID: 25576866
- Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D, He D, Ried T, Symer DE, Gillison ML. 2014. Genome-Wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Research* **24**:185–199. DOI: <https://doi.org/10.1101/gr.164806.113>, PMID: 24201445
- Alexiev BA, Randhawa P, Vazquez Martul E, Zeng G, Luo C, Ramos E, Drachenberg CB, Papadimitriou JC. 2013. Bk virus-associated urinary bladder carcinoma in transplant recipients: report of 2 cases, review of the

- literature, and proposed pathogenetic model. *Human Pathology* **44**:908–917. DOI: <https://doi.org/10.1016/j.humpath.2012.09.019>, PMID: 23317548
- Baker SC**, Mason AS, Slip RG, Skinner KT, Macdonald A, Masood O, Harris RS, Fenton TR, Periyasamy M, Ali S, Southgate J. 2022. Induction of APOBEC3-mediated genomic damage in urothelium implicates BK polyomavirus (bkpyv) as a Hit-and-run driver for bladder cancer. *Oncogene* **41**:2139–2151. DOI: <https://doi.org/10.1038/s41388-022-02235-8>, PMID: 35194151
- Bankevich A**, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**:455–477. DOI: <https://doi.org/10.1089/cmb.2012.0021>, PMID: 22506599
- Barbosa MS**, Edmonds C, Fisher C, Schiller JT, Lowy DR, Vousden KH. 1990. 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* **9**:153–160. DOI: <https://doi.org/10.1002/j.1460-2075.1990.tb08091.x>, PMID: 2153075
- Bianco JN**, Bergoglio V, Lin Y-L, Pillaire M-J, Schmitz A-L, Gilhodes J, Lusque A, Mazières J, Lacroix-Triki M, Roumeliotis TI, Choudhary J, Moreaux J, Hoffmann J-S, Tourrière H, Pasero P. 2019. Overexpression of Claspin and timeless protects cancer cells from replication stress in a checkpoint-independent manner. *Nature Communications* **10**:910. DOI: <https://doi.org/10.1038/s41467-019-08886-8>, PMID: 30796221
- Blatter JA**, Sweet SC, Conrad C, Danziger-Isakov LA, Faro A, Goldfarb SB, Hayes D, Melicoff E, Schecter M, Storch G, Visner GA, Williams NM, Wang D. 2018. Anellovirus loads are associated with outcomes in pediatric lung transplantation. *Pediatric Transplantation* **22**:13069. DOI: <https://doi.org/10.1111/ptr.13069>, PMID: 29082660
- Burns MB**, Temiz NA, Harris RS. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics* **45**:977–983. DOI: <https://doi.org/10.1038/ng.2701>, PMID: 23852168
- Bushmanova E**, Antipov D, Lapidus A, Pribelski AD. 2019. RnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience* **8**:giz100. DOI: <https://doi.org/10.1093/gigascience/giz100>, PMID: 31494669
- Caller LG**, Davies CTR, Antrobus R, Lehner PJ, Weekes MP, Crump CM. 2019. Temporal proteomic analysis of BK polyomavirus infection reveals virus-induced G<sub>2</sub> arrest and highly effective evasion of innate immune sensing. *Journal of Virology* **93**:e00595-19. DOI: <https://doi.org/10.1128/JVI.00595-19>, PMID: 31142673
- Cancer T**, Atlas G, Weinstein JN, Akbani R, Broom BM, Wang W. 2014. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**:315–322. DOI: <https://doi.org/10.1038/nature12965>, PMID: 24476821
- Cantalupo PG**, Katz JP, Pipas JM. 2018. Viral sequences in human cancer. *Virology* **513**:208–216. DOI: <https://doi.org/10.1016/j.virol.2017.10.017>, PMID: 29107929
- Chang Y**, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**:1865–1869. DOI: <https://doi.org/10.1126/science.7997879>, PMID: 7997879
- Chen EY**, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**:128. DOI: <https://doi.org/10.1186/1471-2105-14-128>, PMID: 23586463
- Chen Z**, Wen W, Bao J, Kuhs KL, Cai Q, Long J, Shu X-O, Zheng W, Guo X. 2019. Integrative genomic analyses of APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes, and immunogenicity in multiple cancer types. *BMC Medical Genomics* **12**:131. DOI: <https://doi.org/10.1186/s12920-019-0579-3>, PMID: 31533728
- Crork T**, Tidy JA, Vousden KH. 1991. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. *Cell* **67**:547–556. DOI: [https://doi.org/10.1016/0092-8674\(91\)90529-8](https://doi.org/10.1016/0092-8674(91)90529-8), PMID: 1657399
- Cuenca AG**, Rosales I, Lee RJ, Wu C-L, Colvin R, Feldman AS, Efstathiou JA, Toloff-Rubin N, Elias N. 2020. Resolution of a high grade and metastatic BK polyomavirus-associated urothelial cell carcinoma following radical allograft nephroureterectomy and immune checkpoint treatment: A case report. *Transplantation Proceedings* **52**:2720–2725. DOI: <https://doi.org/10.1016/j.transproceed.2020.06.012>, PMID: 32741665
- Czech-Sioli M**, Günther T, Therre M, Spohn M, Indenbirken D, Theiss J, Riethdorf S, Qi M, Alawi M, Wülbeck C, Fernandez-Cuesta I, Esmek F, Becker JC, Grundhoff A, Fischer N. 2020. High-resolution analysis of merkel cell polyomavirus in merkel cell carcinoma reveals distinct integration patterns and suggests NHEJ and MMBIR as underlying mechanisms. *PLOS Pathogens* **16**:e1008562. DOI: <https://doi.org/10.1371/journal.ppat.1008562>, PMID: 32833988
- Dalianis T**, Hirsch HH. 2013. Human polyomaviruses in disease and cancer. *Virology* **437**:63–72. DOI: <https://doi.org/10.1016/j.virol.2012.12.015>, PMID: 23357733
- D'Arcy ME**, Castenson D, Lynch CF, Kahn AR, Morton LM, Shiels MS, Pfeiffer RM, Engels EA. 2021. Risk of rare cancers among solid organ transplant recipients. *Journal of the National Cancer Institute* **113**:199–207. DOI: <https://doi.org/10.1093/jnci/djaa078>, PMID: 32462187
- DeCaprio JA**. 2014. Human papillomavirus type 16 E7 perturbs DREAM to promote cellular proliferation and mitotic gene expression. *Oncogene* **33**:4036–4038. DOI: <https://doi.org/10.1038/onc.2013.449>, PMID: 24166507
- de Kanter JK**, Peci F, Bertrums E, Rosendahl Huber A, van Leeuwen A, van Roosmalen MJ, Manders F, Verheul M, Oka R, Brandsma AM, Bierings M, Belderbos M, van Boxtel R. 2021. Antiviral treatment causes a



- unique mutational signature in cancers of transplantation recipients. *Cell Stem Cell* **28**:1726–1739. DOI: <https://doi.org/10.1016/j.stem.2021.07.012>, PMID: 34496298
- Dell’Oste V**, Azzimonti B, De Andrea M, Mondini M, Zavattaro E, Leigheb G, Weissenborn SJ, Pfister H, Michael KM, Waterboer T, Pawlita M, Amantea A, Landolfo S, Gariglio M. 2009. High beta-HPV DNA loads and strong seroreactivity are present in epidermodysplasia verruciformis. *The Journal of Investigative Dermatology* **129**:1026–1034. DOI: <https://doi.org/10.1038/jid.2008.317>, PMID: 18923444
- de Martel C**, Georges D, Bray F, Ferlay J, Clifford GM. 2020. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *The Lancet. Global Health* **8**:e180–e190. DOI: [https://doi.org/10.1016/S2214-109X\(19\)30488-7](https://doi.org/10.1016/S2214-109X(19)30488-7), PMID: 31862245
- De Vlamincq I**, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, Okamoto J, Snyder TM, Cornfield DN, Nicolls MR, Weill D, Bernstein D, Valantine HA, Quake SR. 2013. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* **155**:1178–1187. DOI: <https://doi.org/10.1016/j.cell.2013.10.034>, PMID: 24267896
- Dobin A**, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. Star: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: <https://doi.org/10.1093/bioinformatics/bts635>, PMID: 23104886
- Dooley KE**, Warburton A, McBride AA. 2016. Tandemly integrated HPV16 can form a BRD4-dependent super-enhancer-like element that drives transcription of viral oncogenes. *MBio* **7**:e01446-16. DOI: <https://doi.org/10.1128/mBio.01446-16>, PMID: 27624132
- Fang LT**, Afshar PT, Chhibber A, Mohiyuddin M, Fan Y, Mu JC, Gibeling G, Barr S, Asadi NB, Gerstein MB, Koboldt DC, Wang W, Wong WH, Lam HYK. 2015. An ensemble approach to accurately detect somatic mutations using somaticseq. *Genome Biology* **16**:197. DOI: <https://doi.org/10.1186/s13059-015-0758-2>, PMID: 26381235
- Feng H**, Shuda M, Chang Y, Moore PS. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* **319**:1096–1100. DOI: <https://doi.org/10.1126/science.1152586>, PMID: 18202256
- Fischer M**, Müller GA. 2017. Cell cycle transcription control: DREAM/muvb and Rb-E2F complexes. *Critical Reviews in Biochemistry and Molecular Biology* **52**:638–662. DOI: <https://doi.org/10.1080/10409238.2017.1360836>, PMID: 28799433
- Fu F**, Deng W, Yu S, Liu Y, Yu L, Liu R, Lang J, Geng D, Geng J, Li J, Huang G, Luo M, Xiong F, Wu C-L, Miao Y. 2018. Occurrence and regression of BK polyomavirus associated carcinoma: a clinical and next-generation sequencing study. *Clinical Science* **132**:1753–1763. DOI: <https://doi.org/10.1042/CS20180443>, PMID: 30026258
- Graham SV**. 2017. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clinical Science* **131**:2201–2221. DOI: <https://doi.org/10.1042/CS20160786>, PMID: 28798073
- Guma S**, Maglantay R, Lau R, Wiczorek R, Melamed J, Deng F-M, Zhou M, Makarov D, Lee P, Pincus MR, Pei Z-H. 2016. Papillary urothelial carcinoma with squamous differentiation in association with human papilloma virus: case report and literature review. *American Journal of Clinical and Experimental Urology* **4**:12–16 PMID: 27069958.
- Gupta G**, Kuppachi S, Kalil RS, Buck CB, Lynch CF, Engels EA. 2018. Treatment for presumed BK polyomavirus nephropathy and risk of urinary tract cancers among kidney transplant recipients in the United States. *American Journal of Transplantation* **18**:245–252. DOI: <https://doi.org/10.1111/ajt.14530>, PMID: 28980390
- Harris KF**, Christensen JB, Imperiale MJ. 1996. Bk virus large T antigen: interactions with the retinoblastoma family of tumor suppressor proteins and effects on cellular growth control. *Journal of Virology* **70**:2378–2386. DOI: <https://doi.org/10.1128/JVI.70.4.2378-2386.1996>, PMID: 8642665
- Jeannot E**, Harlé A, Holmes A, Sastre-Garau X. 2018. Nuclear factor I X is a recurrent target for HPV16 insertions in anal carcinomas. *Genes, Chromosomes & Cancer* **57**:638–644. DOI: <https://doi.org/10.1002/gcc.22675>, PMID: 30264502
- Jin Y**, Zhou Y, Deng W, Wang Y, Lee RJ, Liu Y, Elias N, Hu Y, Luo M-H, Liu R, Guan B, Geng J, Xu J, Ma J, Zhou J, Liu N, Blute ML, Colvin RB, Wu C-L, Miao Y. 2021. Genome-Wide profiling of BK polyomavirus integration in bladder cancer of kidney transplant recipients reveals mechanisms of the integration at the nucleotide level. *Oncogene* **40**:46–54. DOI: <https://doi.org/10.1038/s41388-020-01502-w>, PMID: 33051598
- Kassem A**, Schöpflin A, Diaz C, Weyers W, Stickeler E, Werner M, Zur Hausen A. 2008. Frequent detection of Merkel cell polyomavirus in human Merkel cell carcinomas and identification of a unique deletion in the VP1 gene. *Cancer Research* **68**:5009–5013. DOI: <https://doi.org/10.1158/0008-5472.CAN-08-0949>, PMID: 18593898
- Kenan DJ**, Mieczkowski PA, Burger-Calderon R, Singh HK, Nickeleit V. 2015. The oncogenic potential of BK-polyomavirus is linked to viral integration into the human genome. *The Journal of Pathology* **237**:379–389. DOI: <https://doi.org/10.1002/path.4584>, PMID: 26172456
- Kenan DJ**, Mieczkowski PA, Latulippe E, Côté I, Singh HK, Nickeleit V. 2017. Bk polyomavirus genomic integration and large T antigen expression: evolving paradigms in human oncogenesis. *American Journal of Transplantation* **17**:1674–1680. DOI: <https://doi.org/10.1111/ajt.14191>, PMID: 28039910
- Koboldt DC**, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* **22**:568–576. DOI: <https://doi.org/10.1101/gr.129684.111>, PMID: 22300766
- Kuleshov MV**, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma’ayan A. 2016. Enrichr: a comprehensive gene

- set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44**:W90–W97. DOI: <https://doi.org/10.1093/nar/gkw377>, PMID: 27141961
- Langmead B**, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**:357–359. DOI: <https://doi.org/10.1038/nmeth.1923>, PMID: 22388286
- Li D**, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* **31**:1674–1676. DOI: <https://doi.org/10.1093/bioinformatics/btv033>, PMID: 25609793
- Li Y-J**, Wu H-H, Chen C-H, Wang H-H, Chiang Y-J, Hsu H-H, Pang S-T, Wang RYL, Tian Y-C. 2021. High incidence and early onset of urinary tract cancers in patients with BK polyomavirus associated nephropathy. *Viruses* **13**:476. DOI: <https://doi.org/10.3390/v13030476>, PMID: 33799453
- Llewellyn MA**, Gordon NS, Abbotts B, James ND, Zeegers MP, Cheng KK, Macdonald A, Roberts S, Parish JL, Ward DG, Bryan RT. 2018. Defining the frequency of human papillomavirus and polyomavirus infection in urothelial bladder tumours. *Scientific Reports* **8**:11290. DOI: <https://doi.org/10.1038/s41598-018-29438-y>, PMID: 30050097
- Love MI**, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with deseq2. *Genome Biology* **15**:550. DOI: <https://doi.org/10.1186/s13059-014-0550-8>, PMID: 25516281
- Martí-Carreras J**, Mineeva-Sangwo O, Topalis D, Snoeck R, Andrei G, Maes P. 2020. BKTyper: free online tool for polyoma BK virus VP1 and NCCR typing. *Viruses* **12**:837. DOI: <https://doi.org/10.3390/v12080837>, PMID: 32751885
- Meier RPH**, Muller YD, Dietrich P-Y, Tille J-C, Nikolaev S, Sartori A, Labidi-Galy I, Hernandez T, Kaur A, Hirsch HH, McKee TA, Toso C, Villard J, Berney T. 2021. Immunologic clearance of a BK virus-associated metastatic renal allograft carcinoma. *Transplantation* **105**:423–429. DOI: <https://doi.org/10.1097/TP.00000000000003193>, PMID: 32091486
- Michel A**, Kopp-Schneider A, Zentgraf H, Gruber AD, de Villiers E-M. 2006. 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. *Journal of Virology* **80**:11153–11164. DOI: <https://doi.org/10.1128/JVI.00954-06>, PMID: 16971438
- Mirabello L**, Yeager M, Yu K, Clifford GM, Xiao Y, Zhu B, Cullen M, Boland JF, Wentzensen N, Nelson CW, Raine-Bennett T, Chen Z, Bass S, Song L, Yang Q, Steinberg M, Burdett L, Dean M, Roberson D, Mitchell J, et al. 2017. Hpv16 E7 genetic conservation is critical to carcinogenesis. *Cell* **170**:1164–1174. DOI: <https://doi.org/10.1016/j.cell.2017.08.001>, PMID: 28886384
- Moens U**, Macdonald A. 2019. Effect of the large and small T-antigens of human polyomaviruses on signaling pathways. *International Journal of Molecular Sciences* **20**:3914. DOI: <https://doi.org/10.3390/ijms20163914>, PMID: 31408949
- Mori S**, Takeuchi T, Ishii Y, Yugawa T, Kiyono T, Nishina H, Kukimoto I. 2017. Human papillomavirus 16 E6 upregulates APOBEC3B via the Tead transcription factor. *Journal of Virology* **91**:e02413-16. DOI: <https://doi.org/10.1128/JVI.02413-16>, PMID: 28077648
- Müller DC**, Rämö M, Naegele K, Ribi S, Wetterauer C, Perrina V, Quagliata L, Vlajnic T, Ruiz C, Balitzki B, Grobholz R, Gosert R, Ajuh ET, Hirsch HH, Bubendorf L, Rentsch CA. 2018. Donor-Derived, metastatic urothelial cancer after kidney transplantation associated with a potentially oncogenic BK polyomavirus. *The Journal of Pathology* **244**:265–270. DOI: <https://doi.org/10.1002/path.5012>, PMID: 29205775
- Nassar AH**, Umeton R, Kim J, Lundgren K, Harshman L, Van Allen EM, Preston M, Dong F, Bellmunt J, Mouw KW, Choueiri TK, Sonpavde G, Kwiatkowski DJ. 2019. Mutational analysis of 472 urothelial carcinoma across grades and anatomic sites. *Clinical Cancer Research* **25**:2458–2470. DOI: <https://doi.org/10.1158/1078-0432.CCR-18-3147>, PMID: 30593515
- Nguyen DTT**, Richter D, Michel G, Mitschka S, Kolanus W, Cuevas E, Wulczyn FG. 2017. The ubiquitin ligase LIN41/TRIM71 targets p53 to antagonize cell death and differentiation pathways during stem cell differentiation. *Cell Death and Differentiation* **24**:1063–1078. DOI: <https://doi.org/10.1038/cdd.2017.54>, PMID: 28430184
- Nickeleit V**, Singh HK, Kenan DJ, Mieczkowski PA. 2018. The two-faced nature of BK polyomavirus: lytic infection or non-lytic large-T-positive carcinoma. *The Journal of Pathology* **246**:7–11. DOI: <https://doi.org/10.1002/path.5127>, PMID: 29931826
- Papadimitriou JC**, Randhawa P, Rinaldo CH, Drachenberg CB, Alexiev B, Hirsch HH. 2016. Bk polyomavirus infection and renourinary tumorigenesis. *American Journal of Transplantation* **16**:398–406. DOI: <https://doi.org/10.1111/ajt.13550>, PMID: 26731714
- Peretti A**, Geoghegan EM, Pastrana DV, Smola S, Feld P, Sauter M, Lohse S, Ramesh M, Lim ES, Wang D, Borgogna C, FitzGerald PC, Bliskovsky V, Starrett GJ, Law EK, Harris RS, Killian JK, Zhu J, Pineda M, Meltzer PS, et al. 2018. Characterization of BK polyomaviruses from kidney transplant recipients suggests a role for APOBEC3 in driving in-host virus evolution. *Cell Host & Microbe* **23**:628–635. DOI: <https://doi.org/10.1016/j.chom.2018.04.005>, PMID: 29746834
- Poon SL**, Huang MN, Choo Y, McPherson JR, Yu W, Heng HL, Gan A, Myint SS, Siew EY, Ler LD, Ng LG, Weng W-H, Chuang C-K, Yuen JS, Pang S-T, Tan P, Teh BT, Rozen SG. 2015. Mutation signatures implicate aristolochic acid in bladder cancer development. *Genome Medicine* **7**:38. DOI: <https://doi.org/10.1186/s13073-015-0161-3>, PMID: 26015808
- Poulain F**, Lejeune N, Willemart K, Gillet NA. 2020. Footprint of the host restriction factors APOBEC3 on the genome of human viruses. *PLOS Pathogens* **16**:e1008718. DOI: <https://doi.org/10.1371/journal.ppat.1008718>, PMID: 32797103

- Querido S**, Fernandes I, Weigert A, Casimiro S, Albuquerque C, Ramos S, Adragão T, Luz I, Paixão P, Chasqueira M, Santos M, Machado D. 2020. High-grade urothelial carcinoma in a kidney transplant recipient after JC virus nephropathy: the first evidence of JC virus as a potential oncovirus in bladder cancer. *American Journal of Transplantation* **20**:1188–1191. DOI: <https://doi.org/10.1111/ajt.15663>, PMID: 31654479
- Robbe P**, Popitsch N, Knight SJL, Antoniou P, Becq J, He M, Kanapin A, Samsonova A, Vavoulis DV, Ross MT, Kingsbury Z, Cabel M, Ramos SDC, Page S, Dreau H, Ridout K, Jones LJ, Tuff-Lacey A, Henderson S, Mason J, et al. 2018. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 genomes project. *Genetics in Medicine* **20**:1196–1205. DOI: <https://doi.org/10.1038/gim.2017.241>, PMID: 29388947
- Roberts SA**, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, Getz G, Gordenin DA. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics* **45**:970–976. DOI: <https://doi.org/10.1038/ng.2702>, PMID: 23852170
- Robertson AG**, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, Castro MAA, Gibb EA, Kanchi RS, Gordenin DA, Shukla SA, Sanchez-Vega F, Hansel DE, Czerniak BA, Reuter VE, Su X, et al. 2017. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**:540–556. DOI: <https://doi.org/10.1016/j.cell.2017.09.007>, PMID: 28988769
- Rosty C**, Sheffer M, Tsafirir D, Stransky N, Tsafirir I, Peter M, de Crémoux P, de La Rochefordière A, Salmon R, Dorval T, Thierry JP, Couturier J, Radvanyi F, Domany E, Sastre-Garau X. 2005. Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. *Oncogene* **24**:7094–7104. DOI: <https://doi.org/10.1038/sj.onc.1208854>, PMID: 16007141
- Sadasivam S**, Duan S, DeCaprio JA. 2012. The mvb complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes & Development* **26**:474–489. DOI: <https://doi.org/10.1101/gad.181933.111>, PMID: 22391450
- Schade AE**, Oser MG, Nicholson HE, DeCaprio JA. 2019. Cyclin D-Cdk4 relieves cooperative repression of proliferation and cell cycle gene expression by DREAM and Rb. *Oncogene* **38**:4962–4976. DOI: <https://doi.org/10.1038/s41388-019-0767-9>, PMID: 30833638
- Schneidewind L**, Neumann T, Plis A, Brückmann S, Keiser M, Krüger W, Schmidt CA. 2020. Novel 3D organotypic urothelial cell culture model for identification of new therapeutic approaches in urological infections. *Journal of Clinical Virology* **124**:104283. DOI: <https://doi.org/10.1016/j.jcv.2020.104283>, PMID: 31986366
- Serebrenik AA**, Starrett GJ, Leenen S, Jarvis MC, Shaban NM, Salamango DJ, Nilsen H, Brown WL, Harris RS. 2019. The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. *PNAS* **116**:22158–22163. DOI: <https://doi.org/10.1073/pnas.1904024116>, PMID: 31611371
- Shuda M**, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y. 2008. T antigen mutations are a human tumor-specific signature for merkel cell polyomavirus. *PNAS* **105**:16272–16277. DOI: <https://doi.org/10.1073/pnas.0806526105>, PMID: 18812503
- Sirohi D**, Vaske C, Sanborn Z, Smith SC, Don MD, Lindsey KG, Federman S, Vankalakunti M, Koo J, Bose S, Peralta-Venturina M de, Ziffle J, Grenert JP, Miller S, Chiu C, Amin MB, Simko JP, Stohr BA, Luthringer DJ. 2018. Polyoma virus-associated carcinomas of the urologic tract: a clinicopathologic and molecular study. *Modern Pathology* **31**:1429–1441. DOI: <https://doi.org/10.1038/s41379-018-0065-z>, PMID: 29765141
- Solis M**, Velay A, Porcher R, Domingo-Calap P, Soulier E, Joly M, Meddeb M, Kack-Kack W, Moulin B, Bahram S, Stoll-Keller F, Barth H, Caillard S, Fafi-Kremer S. 2018. Neutralizing antibody-mediated response and risk of BK virus-associated nephropathy. *Journal of the American Society of Nephrology* **29**:326–334. DOI: <https://doi.org/10.1681/ASN.2017050532>, PMID: 29042457
- Starrett GJ**, Marcelus C, Cantalupo PG, Katz JP, Cheng J, Akagi K, Thakuria M, Rabinowits G, Wang LC, Symer DE, Pipas JM, Harris RS, DeCaprio JA. 2017. Merkel cell polyomavirus exhibits dominant control of the tumor genome and transcriptome in virus-associated Merkel cell carcinoma. *MBio* **8**:e02079-16. DOI: <https://doi.org/10.1128/mBio.02079-16>, PMID: 28049147
- Starrett GJ**, Serebrenik AA, Roelofs PA, McCann JL, Verhalen B, Jarvis MC, Stewart TA, Law EK, Krupp A, Jiang M, Martens JWM, Cahir-McFarland E, Span PN, Harris RS. 2019. Polyomavirus T antigen induces APOBEC3B expression using an LXCXE-dependent and TP53-independent mechanism. *MBio* **10**:e02690-18. DOI: <https://doi.org/10.1128/mBio.02690-18>, PMID: 30723127
- Starrett GJ**. 2020. Oncovirus tools. 99c8707. GitHub. [https://github.com/gstarrett/oncovirus\\_tools](https://github.com/gstarrett/oncovirus_tools)
- Starrett GJ**, Thakuria M, Chen T, Marcelus C, Cheng J, Nomburg J, Thorner AR, Slevin MK, Powers W, Burns RT, Perry C, Piris A, Kuo FC, Rabinowits G, Giobbie-Hurder A, MacConaill LE, DeCaprio JA. 2020. Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma. *Genome Medicine* **12**:30. DOI: <https://doi.org/10.1186/s13073-020-00727-4>, PMID: 32188490
- Su X**, Lu X, Bazai SK, Compérat E, Mouawad R, Yao H, Rouprêt M, Spano J-P, Khayat D, Davidson I, Tannir NN, Yan F, Malouf GG. 2021. Comprehensive integrative profiling of upper tract urothelial carcinomas. *Genome Biology* **22**:7. DOI: <https://doi.org/10.1186/s13059-020-02230-w>, PMID: 33397444
- Swanton C**, McGranahan N, Starrett GJ, Harris RS. 2015. Apobec enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discovery* **5**:704–712. DOI: <https://doi.org/10.1158/2159-8290.CD-15-0344>, PMID: 26091828
- Tisza MJ**, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, Pang Y-YS, Krishnamurthy SR, Pesavento PA, McDermott DH, Murphy PM, Whited JL, Miller B, Brenchley J, Rosshart SP, Rehmann B, Doorbar J, Ta'ala BA, Pletnikova O, Troncoso JC, et al. 2020. Discovery of several thousand highly diverse circular DNA viruses. *eLife* **9**:e51971. DOI: <https://doi.org/10.7554/eLife.51971>, PMID: 32014111

- Tisza M.** 2021. Cenote-taker 2. 2.1.2. GitHub. <https://github.com/mtisza1/Cenote-Taker2>
- Tisza MJ,** Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. 2021. Cenote-taker 2 democratizes virus discovery and sequence annotation. *Virus Evolution* **7**:veaa100. DOI: <https://doi.org/10.1093/ve/veaa100>, PMID: 33505708
- Toptan T,** Yousem SA, Ho J, Matsushima Y, Stabile LP, Fernández-Figueroa M-T, Bhargava R, Ryo A, Moore PS, Chang Y. 2016. Survey for human polyomaviruses in cancer. *JCI Insight* **1**:e85562. DOI: <https://doi.org/10.1172/jci.insight.85562>, PMID: 27034991
- Tye S,** Ronson GE, Morris JR. 2021. A fork in the road: where homologous recombination and stalled replication fork protection part ways. *Seminars in Cell & Developmental Biology* **113**:14–26. DOI: <https://doi.org/10.1016/j.semcdb.2020.07.004>, PMID: 32653304
- Vajdic CM,** van Leeuwen MT. 2009. Cancer incidence and risk factors after solid organ transplantation. *International Journal of Cancer* **125**:1747–1754. DOI: <https://doi.org/10.1002/ijc.24439>, PMID: 19444916
- Van Doorslaer K,** Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, Sun Q, Kaur R, Huyen Y, McBride AA. 2017. The papillomavirus episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Research* **45**:D499–D506. DOI: <https://doi.org/10.1093/nar/gkw879>, PMID: 28053164
- Verhalen B,** Starrett GJ, Harris RS, Jiang M. 2016. Functional upregulation of the DNA cytosine deaminase APOBEC3B by polyomaviruses. *Journal of Virology* **90**:6379–6386. DOI: <https://doi.org/10.1128/JVI.00771-16>, PMID: 27147740
- Viarisio D,** Müller-Decker K, Zanna P, Kloz U, Aengeneyndt B, Accardi R, Flechtenmacher C, Gissmann L, Tommasino M. 2016. Novel  $\beta$ -HPV49 transgenic mouse model of upper digestive tract cancer. *Cancer Research* **76**:4216–4225. DOI: <https://doi.org/10.1158/0008-5472.CAN-16-0370>
- Viarisio D,** Müller-Decker K, Accardi R, Robitaille A, Dürst M, Beer K, Jansen L, Flechtenmacher C, Bozza M, Harbottle R, Voegelé C, Ardin M, Zavadil J, Caldeira S, Gissmann L, Tommasino M. 2018. Beta HPV38 oncoproteins act with a hit-and-run mechanism in ultraviolet radiation-induced skin carcinogenesis in mice. *PLoS Pathogens* **14**:e1006783. DOI: <https://doi.org/10.1371/journal.ppat.1006783>, PMID: 29324843
- Vieira VC,** Leonard B, White EA, Starrett GJ, Temiz NA, Lorenz LD, Lee D, Soares MA, Lambert PF, Howley PM, Harris RS. 2014. Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *MBio* **5**:e02234-14. DOI: <https://doi.org/10.1128/mBio.02234-14>, PMID: 25538195
- Wallez Y,** Riedl SJ, Pasquale EB. 2014. Association of the breast cancer antiestrogen resistance protein 1 (BCAR1) and BCAR3 scaffolding proteins in cell signaling and antiestrogen resistance. *The Journal of Biological Chemistry* **289**:10431–10444. DOI: <https://doi.org/10.1074/jbc.M113.541839>, PMID: 24584939
- Wang Y,** Liu Y, Deng W, Fu F, Yan S, Yang H, Liu R, Geng J, Xu J, Wu Y, Ma J, Zhou J, Liu N, Jin Y, Xia R, Elias N, Lee RJ, Feldman AS, Blute ML, Colvin RB, et al. 2020. Viral integration in BK polyomavirus-associated urothelial carcinoma in renal transplant recipients: multistage carcinogenesis revealed by next-generation virome capture sequencing. *Oncogene* **39**:5734–5742. DOI: <https://doi.org/10.1038/s41388-020-01398-6>
- Warburton A,** Redmond CJ, Dooley KE, Fu H, Gillison ML, Akagi K, Symer DE, Aladjem MI, McBride AA. 2018. HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression. *PLOS Genetics* **14**:e1007179. DOI: <https://doi.org/10.1371/journal.pgen.1007179>, PMID: 29364907
- Warren CJ,** Van Doorslaer K, Pandey A, Espinosa JM, Pyeon D. 2015a. Role of the host restriction factor APOBEC3 on papillomavirus evolution. *Virus Evolution* **1**:vev015. DOI: <https://doi.org/10.1093/ve/vev015>, PMID: 27570633
- Warren CJ,** Xu T, Guo K, Griffin LM, Westrich JA, Lee D, Lambert PF, Santiago ML, Pyeon D. 2015b. Apobec3A functions as a restriction factor of human papillomavirus. *Journal of Virology* **89**:688–702. DOI: <https://doi.org/10.1128/JVI.02383-14>, PMID: 25355878
- Warren C.J,** Westrich JA, Doorslaer KV, Pyeon D. 2017. Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. *Viruses* **9**:233. DOI: <https://doi.org/10.3390/v9080233>, PMID: 28825669
- Wilm A,** Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research* **40**:11189–11201. DOI: <https://doi.org/10.1093/nar/gks918>, PMID: 23066108
- Wilson AL,** Schrecengost RS, Guerrero MS, Thomas KS, Bouton AH. 2013. Breast cancer antiestrogen resistance 3 (BCAR3) promotes cell motility by regulating actin cytoskeletal and adhesion remodeling in invasive breast cancer cells. *PLOS ONE* **8**:e65678. DOI: <https://doi.org/10.1371/journal.pone.0065678>, PMID: 23762409
- Zapatka M,** Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, Desai N, Sültmann H, Moch H, PCAWG Pathogens, Cooper CS, Eils R, Ferretti V, Lichter P, PCAWG Consortium. 2020. The landscape of viral associations in human cancers. *Nature Genetics* **52**:320–330. DOI: <https://doi.org/10.1038/s41588-019-0558-9>, PMID: 32025001
- Zhang J,** Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, Seth S, Chow C-W, Cao Y, Gumbs C, Gold KA, Kalhor N, Little L, Mahadeshwar H, Moran C, Protopopov A, Sun H, Tang J, Wu X, Ye Y, et al. 2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**:256–259. DOI: <https://doi.org/10.1126/science.1256930>, PMID: 25301631