Tumor-specific changes in Kaposi sarcoma-associated herpesvirus genomes in Ugandan adults with Kaposi sarcoma

Jan Clement Santiago[1]*, Jason Goldman[2,3]*, Hong Zhao[1], Alec Pankow[1], Fred Okuku[4], Michael Schmitt[2], Lennie Chen[1], C. Alexander Hill[1], Corey Casper[2,3^], Warren T. Phipps[2,3] and James I. Mullins[1,2,6,7±]

University of Washington, Departments of Microbiology[1], Medicine[2], Global Health[6] and Laboratory Medicine,[7] Seattle, WA US

[3] Fred Hutchinson Cancer Research Center, Seattle, WA US

[4] Uganda Cancer Institute, Kampala, UG

^ Present address: Infectious Disease Research Institute, Seattle, WA US

± Corresponding author:

E-mail: jmullins@uw.edu (JM)

* These authors contributed equally to this work

Abstract

Intra-host evolved tumor virus variants have provided insights into the risk, pathogenesis and treatment responses of associated cancers. However, the intra-host variability of Kaposi sarcoma-associated herpesvirus (KSHV), the etiologic agent of Kaposi sarcoma (KS), has not been explored at the whole viral genome level. An accurate and detailed description of KSHV intra-host diversity in whole KSHV genomes from matching tumors and oral swabs from Ugandan adults with HIV-associated KS was obtained by deep, short read sequencing, using duplex unique molecular identifiers (dUMI) – random double-stranded oligonucleotides that barcode individual DNA molecules before library amplification. This allowed suppression of PCR and sequencing errors down to ~10^{-9}/base. KSHV genomes were assembled *de novo*, and identified rearrangements were confirmed by PCR. 131-kb KSHV genome sequences, excluding major repeat regions and averaging 2.3 x 10^4 reads/base, were successfully obtained from 23 specimens from 9 individuals, including 7 tumor-oral pairs. Sampling more than 100 viral genomes in at least one specimen per individual showed that KSHV genomes were virtually homogeneous within samples and within individuals at the point mutational level. Heterogeneity, if present, was due to point mutations and genomic rearrangements in tumors. In 2 individuals, the same mutations were found in distinct KS tumors. The K8.1 gene was inactivated in tumors from 3 individuals, and all KSHV genomic aberrations retained the region surrounding the first major internal repeat (IR1). These findings suggest that lytic gene alterations may contribute to KS tumorigenesis or persistence.

Author summary

Kaposi sarcoma (KS) is a leading cancer in sub-Saharan Africa and in those with HIV co-infection. Infection by Kaposi sarcoma-associated herpesvirus (KSHV) is necessary for KS, yet why only few KSHV infections develop into KS is largely unknown. While strain differences or mutations in other tumor viruses

are known to affect the risk and progression of their associated cancers, whether KSHV genetic variation is important to the natural history of KS is unclear. Most studies of KSHV diversity have characterized only ~4% of its 165-kb genome and may have been impacted by PCR or cloning artifacts. Here, we performed highly sensitive, single-molecule sequencing of whole KSHV genomes in paired KS tumors and oral swabs from 9 individuals with KS. We found that KSHV genomes were virtually identical within individuals, with no evidence of quasispecies formation nor multistrain infection. However, KSHV genome aberrations and inactivating mutations appeared to be a common, tumor-associated phenomenon, with some mutations shared by distinct tumors within an individual. Certain regions of the KSHV genome featured prominently among tumor-associated mutations, suggesting that they are important contributors to the pathogenesis or persistence of KS.

Introduction

Kaposi Sarcoma (KS) is one of the most common cancers of HIV-infected individuals [1,2], with the burden of disease disproportionately borne by people in sub-Saharan Africa [3]. A gamma herpesvirus, Kaposi sarcoma-associated herpesvirus (KSHV), is the etiologic agent of KS and consistently detected in tumor tissues [4,5]. KSHV also can be shed in saliva, thought to be a primary mode of transmission [6–8]. Only a small fraction of KSHV infections progress to KS, and the factors contributing to KS pathogenesis are poorly understood. The development of KS is associated with HIV infection and immunosuppression [9], but others factors, including KSHV genome variation, may contribute to differential outcomes of KSHV infection.

Studies of other human oncogenic viruses reveal that viral genetic variation or *de novo* mutations may be important to their pathogenicity, as is the case for cancers associated with human papilloma viruses and Merkel cell polyomavirus [10]. Epstein Bar virus (EBV), another gamma-herpesvirus like KSHV,

is associated with a variety of neoplasms. In EBV infections, intra-host evolved viruses may have a role in pathogenesis of associated cancers [11–13]. Additionally, EBV strains isolated from nasopharyngeal carcinoma (NPC) biopsies have unique genomic [14,15] and phenotypic [16–18] variations compared to other strains and isolates from geographically clustered individuals without cancer [19]. NPC-associated strains were also found to have increased tropism for epithelial- versus B-cells [16,20]. These findings suggest that viral genetic heterogeneity can affect EBV virulence. Whether KSHV genetic variation can similarly influence KS pathogenesis or manifestation is unknown.

KSHV encodes oncogenes that dysregulate cell cycle, cell-to-cell adhesion, inflammation and angiogenesis [21–24], and heterogeneity in these genes might result in differences in KSHV infection outcomes and clinical manifestations. For example, polymorphisms in the microRNA region of KSHV have been correlated with the development of multicentric Castleman disease and KSHV-associated inflammatory cytokine syndrome with and without KS [25]. The K1 gene, the most variable KSHV gene, is conventionally used for KSHV strain subtyping. Subtype A has been associated with more aggressive KS than subtype C [26,27], while subtype B has been associated with better KS prognoses [28]. However, correlations of KSHV genetic subtypes with virulence has not been consistently observed [29–32].

KSHV whole genome sequences provide a far more comprehensive picture of KSHV diversity than the KSHV variable regions alone. Most of the publicly available KSHV genomes have been reported in only the last 5 years [33–37], and they demonstrated that polymorphisms in the ~130-kb KSHV non-repeat genomic region outside of the K1 gene contribute much more to KSHV diversity than the 0.9 kb hyper-variable K1 gene by itself [33,34]. KSHV genomes are, moreover, replete with signatures of recombination [36], potentially complicating disease risk association solely with K1 subtypes. Principal component analysis of 70 KSHV genomes, all from different individuals, failed to reveal strain specificity for the development of KS [36].

Whether KSHV results in significant intra-host diversity is unclear. Some studies examining KSHV in different compartments or multiple clones from a single individual have reported KSHV quasispecies, multi-strain infections [38–42] and intra-host evolution [37,40,43], while only a single persisting strain is typically found within individuals with AIDS-associated KS [44–47]. The existence of KSHV intertypic recombinants [36,44,48–50] indicate that co-infection of divergent KSHV strains must be occurring at least sporadically.

The assessment of intra-host diversity can be easily biased by artifacts introduced during sample preparation and when sequencing from PCR products [45]. Short read, next generation sequencing can have high error rates due to PCR misincorporation, end-repair artifacts, insufficient sequencing depth, and DNA damage from long, repeated high-temperature incubations during PCR and enrichment reactions [51–54]. These methodological limitations can undermine interpretation of observed intra-host KSHV variations.

To accurately assess KSHV intra-host polymorphisms associated with KS tumors and detect minor KSHV sequence variants, we sequenced KSHV genomes using a highly sensitive short-read sequencing method termed “duplex sequencing” [55]. This method incorporates duplex unique molecular identifiers (dUMI), which are double-stranded strings of random base pairs to barcode individual DNA molecules before PCR amplification and RNA-bait enrichment [55]. By utilizing dUMI-consensus reads of each DNA molecule in a sample library, PCR-associated errors are reduced to ~10^{-9}, revealing the original sequence variation within a sample library [55]. In the present study, we report the results of duplex sequencing of whole KSHV genomes in paired tumors and oral swabs of 9 Ugandan adults with HIV-associated KS.

Methods

Study Cohort and Specimen collection.

Specimens were obtained from participants enrolled in the "HIPPOS" Study, an ongoing prospective cohort study, begun in 2012, of KS patients initiating treatment at the Uganda Cancer Institute (UCI) in Kampala, Uganda. This protocol was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board, the Makerere University School of Medicine Research and Ethics Committee (SOMREC), and the Uganda National Council on Science and Technology (UNCST). All participants provided written informed consent.

Participants were eligible for the HIPPOS study if they were >18 years of age with biopsy-proven KS, and ART- and chemotherapy-naïve at enrollment. Participants attended 12 study visits over a one-year period and received treatment for KS consisting of ART and chemotherapy (combination bleomycin and vincristine or paclitaxel) per standard protocols by UCI physicians. At each visit, participants received a detailed physical exam to assess clinical response using the ACTG KS response criteria [56].

Participants provided plasma samples at each visit for KSHV, CD4 and HIV viral load testing, and in addition, provided up to 9 biopsies of KS lesions before, during, and after KS treatment. KS tumor biopsies were obtained using 4mm punch biopsy tools after cleaning the skin with alcohol, and either snap-frozen at the clinic site and stored in liquid nitrogen (LN2) or placed in RNAlater and stored at -80^{0}C. Study clinicians collected swabs of the oral mucosa at each study visit and participants self-collected oral swabs at home for 1 week after the visit after education on the sample collection technique, as has been previously validated by our group in Uganda [57]. Briefly, a Dacron swab is inserted into the mouth and vigorously rubbed along the buccal mucosa, gums, and hard palate. The swab is then placed in 1 mL of filter-sterilized digestion buffer [58] and stored at ambient temperature [59] before being placed at -20^{o}C for storage.

DNA preparation. DNA was extracted from 300µL homogenized tumor lysates using the AllPrep DNA/RNA Mini Kit (QIAGEN, Cat. # 80204) and eluted into 100µL EB Buffer. For oral swab specimens, DNA was extracted from the swab tip eluate using the QIAamp Mini Kit (Qiagen, Cat.# 51304) following the manufacturer's protocol. Purification of DNA from saliva stabilized in RNAprotect® Saliva Reagent (Qiagen) was performed following the manufacturer's protocol with the following modifications: there was no initial pelleting or PBS wash, 20 µL proteinase K was used per 200 µL specimen, and DNA was eluted in 50 µL water. DNA was quantified using a NanoDrop™ 1000 Spectrophotometer (ThermoScientific).

PCR. All PCR preparations were done in a PCR-clean room, except for the addition of control templates. PCR was conducted using the PrimeSTAR GXL kit (Takara, Cat. # R050B) with ThermaStop™ (Thermagenix) added. Cycling conditions were: 98°C for 2 mins; 35 cycles of 98˚C for 10 secs, 60-65˚C (depending on primer) for 15 secs, 68˚C for 1min/kb; 68˚C for 3 mins and then hold at 4˚C. Primer sequences are listed in **S1 Table**.

Copy number quantification. KSHV genome copy numbers were quantified by digital droplet PCR (ddPCR) using the QX200™ Droplet Generator and Reader (Bio-Rad), with ddPCR™ SuperMix for Probes No dUTP (Bio-Rad, Cat. # 186-3024). Primers and probes were designed (**S1 Table**) to detect 4 KSHV-unique genes K2/vIL-6, ORF16/vBCL-2, ORF50/RTA and ORF73/LANA (**Fig 1A**), and the KSHV genome copy numbers reported were the average of the 4 measures. 420 ng BCBL-1 cell line DNA diluted 1:10,000, about 475 genome copies, was used as positive control. 1 ng human genomic DNA (Bioline, Cat. # BIO-35025) was used as negative control, and water as a no template control. Cycling conditions were: 95°C for 10 mins; 40 cycles of: 94°C for 30 secs, 56°C for 30 secs, 60°C for 1 min; one cycle at 98°C 10 for mins, and then

hold at 12°C. The KSHV on-target percent was calculated using the copy number quantification by ddPCR normalized to the total nucleic acid concentration.

UMI-addition and library preparation (Fig 2). To obtain ~500-bp DNA fragments, 10-20 ng/µL of DNA extract in 100 µL chilled TLE buffer (10mM Tris, pH8.0, 0.1mM EDTA) was sheared using a Bioruptor™ (Diagenode) on high power for up to 15 min. Fragment sizes were assessed on 1.5% agarose gels. Sheared DNA was bead-purified using 1.2X volume of Agencourt AMPure XP Beads (Beckman Coulter Cat. # A63880) and eluted in 50 µL water. Library preparation (end repair, A-tailing and adapter-ligation) was performed using the KAPA HyperPrep Library Preparation Kit (Cat. # KR0961/KK8503). Double-stranded DNA adapters contained a random 12-bp dUMI sequence and a defined 5-bp spacer sequence added to Illumina TruSeq adaptor sequences [60] (**S1 Fig**). Subsequently, DNA was bead-purified with 1X volume of beads and eluted in 50 µL water.

DNA libraries were subjected to pre-enrichment amplification with primers mws13 & mws20 (**S1 Table**, **S1 Fig**) and KAPA HiFi Hot Start polymerase. PCR conditions were: 95°C 4 mins; 5-8 cycles of 98°C 20 sec, 60°C 45 sec, 72°C 45 sec; 72°C 3 mins, 4°C hold. If the bead-purified elution from the end repair and adapter step had more than 240 ng total, it was divided into 50 µL PCR reactions of $\leq$240 ng and pooled after amplification. PCR products were then bead-purified as above with 1.2X volume beads and elution in 100 µL water, quantified with Nanodrop, and their sizes assessed using a Bio-analyzer (Agilent DNA 7500) or Qiaxcel (QIAGEN AM420).

Library enrichment & sequencing. Biotinylated RNA baits for enriching KSHV sequences in the library were those designed in [61] and were obtained from Agilent, Inc. (Santa Clara, CA). The design was a 120-bp, 12X tiling of the genome of KSHV isolate GK18 (Genbank ID: AF148805.2). The diversity of the bait library

was further increased by including K1, ORF75, K15, ORF26 and TR sequences of strains JSC-1 (Genbank ID: GQ994935.1), DG1 (Genbank ID: JQ619843.1), BC-1 (Genbank ID: U75698.1), BCBL-1 (Genbank ID: HQ404500.1), Sau3A (Genbank ID: U93872.2), and all Western and African strain sequences in [29,33] (Genbank ID: AF130259, AF130266, AF130267, AF130281, AF130305, AF133039, AF133040, AF133043, AF133044, AF151687, AF171057, AF178780, AF178810, AF220292, AF220293, AY329032, KT271453, KT271454, KT271455, KT271456, KT271457, KT271458, KT271459, KT271460, KT271461, KT271462, KT271463, KT271464, KT271465, KT271466, KT271467, KT271468).

Target enrichment was performed using SureSelect Target Enrichment Kit v1.7 (Agilent) with all suggested volumes reduced by half. DNA hybridized to biotinylated-RNA baits was captured with streptavidin beads (Dynabeads MyOne Streptavidin T1, Invitrogen) and resuspended in 20µL water. The DNA-streptavidin bead mixture was used directly in post-enrichment PCR amplification with primers mws13 and mws21, the latter of which includes a sample index sequence (**S1 Table**, **S1 Fig**). The PCR cycle number ranged from 10-16, with products monitored every 2 to 3 cycles on a TapeStation (Agilent) to ensure correct fragment sizes (~500bp). When over-amplification resulted in library fragment sizes much larger than expected, a single "reconditioning" PCR cycle with fresh reagents was done [62]. PCR products were cleaned using 1.2X volume AMPure XP beads and the eluted DNA library was sequenced using Illumina HiSeq 2500 with 100-bp paired end reads. For some tumor samples with low KSHV copy numbers and all oral swab samples, a second library enrichment was performed.

***De novo* assembly of sample-consensus genomes**. Initially, a sample-consensus KSHV genome (**Fig 2**) was generated *de novo* from paired-end reads of each sample using custom scripts (**S2 Fig**, https://github.com/MullinsLab/HHV8-assembly-SPAdes). At this stage, the first 17-bp from read ends were trimmed to remove dUMI sequences. Next, reads were subjected to windowed quality filtering using

sickle pe [63] with a quality threshold of 30 and a window size 10% of read length. Filtered reads were aligned to a human genome (GRCh38 p12, GenBank GCA_000001405.27) using *bwa mem* [64]. Unmapped reads were used as input into the de novo assembler SPAdes v3.11.1 [65], with the setting *-k 21,35,55,71,81*. This oftentimes yielded 3 to 4 scaffolds that together encompassed the entire 131-kb unique sequence regions of KSHV, bounded by the major repeat regions: Internal Repeat 1 (IR1), Internal Repeat 2 (IR2), LANA central repeat and Terminal Repeats (TR) (**Fig 1A**). Next, all scaffolds over 500 bp were aligned using *bwa mem* to the genome of reference KSHV strain GK18. From the aligned scaffolds a draft genome was generated in Geneious (Biomatters, Ltd) with manual correction as needed. To finish the assembly, GapFiller v1.1 [66] was used, setting *bwa* as the aligner and filtered paired-end reads as the input library. The genomes were annotated in Geneious from the GK18 reference, also adding the T1.4 annotation based on [67]. The major repeat regions were masked with Ns since they were poorly resolved by assembly of short reads that can map to multiple locations within the repeat regions.

Variant identification from dUMI-consensus reads. Paired-end reads, including their dUMI sequence tags, were mapped by *bwa* to sample-consensus genomes (**Fig 2**) using a Makefile adapted from [60] (https://github.com/MullinsLab/Duplex-Sequencing). Briefly, all reads mapping to the same genomic position were collapsed by single strand UMIs (sUMI) to make sUMI-consensus reads (**S2 Fig**). Complementary UMI tags from opposing strands were matched to create dUMI-consensus reads, removing nearly all PCR polymerase misincorporation and chimera artifacts. Nine bases from both read ends were then trimmed to minimize read end artifacts. Discrepancies between mapped dUMI-consensus reads and the sample-consensus genomes were manually inspected and corrected in Geneious as needed. Only the remaining discrepancies were considered to be sequence variants that existed prior to PCR amplification.

All genome and subgenome sequence alignments were done using MAFFT [68] [algorithm FFT-NS-i x1000, scoring matrix 1PAM/k=2], and all phylogenetic trees were inferred using RAxML [69] (-f d, GTR gamma, N=100 starting trees), using a representative KSHV genome from each individual. The NeighborNet phylogenetic network was generated using SplitsTree5, excluding gap sites [70].

Consensus genome sequences were deposited in GenBank (Accession numbers: XXX) with coordinates of rearrangements, when present, indicated.

Integration analysis. Systematic searches for KSHV integration into the human genome were done in two ways. First, each library was searched using local BLASTN against both human and KSHV sequences and then using the Perl script SummonChimera [71] to extract coordinates of potential integration sites. Second, a sample-consensus KSHV genome was appended as an extra chromosome to the human genome reference GRCh38 p12. The appended human genome reference was used to map sUMI consensus reads via Speedseq [72] to generate alignment files with only discordant or split reads. These were input into LUMPY for structural variant analysis [73]. Human chromosomes linked to KSHV sequences were taken to be putative integration sites.

Results

Assessment of the dUMI sequencing protocol with a KSHV infected cell line

As part of the optimization of the dUMI-sequencing protocol, KSHV genome sequences were first obtained from an early passage of BCBL-1, a KSHV-infected PEL cell line [74]. BCBL-1 cells were grown as previously described [75]. After DNA extraction, KSHV sequences corresponded to ~0.16% of the total DNA using a ddPCR assay for ORF73 and T0.7-K12, and normalized by comparison to the human gene

247 POLG. Following a single round of bait capture, the fraction of sequence reads corresponding to KSHV
248 from BCBL-1 DNA extracts (i.e., the "on-target" level), was 15.6%, corresponding to 173-fold enrichment.

249 Sequencing of the BCBL-1 KSHV genome produced a mean coverage of >10,000 reads per base
250 excluding the repeat regions. Collapsing raw reads by identical sUMI to generate sUMI-consensus reads
251 results in a mean of 2,552 sUMI reads per base. When collapsed further into consensus sequences derived
252 from both strands, a mean of 286 dUMI reads per base was obtained that was essentially free of PCR
253 errors (**Table 1; Fig 2B**). Since each dUMI tags a unique DNA molecule before PCR, the number of unique
254 dUMI tags indicates the number of unique viral templates sequenced [55,76]. Using this measure, 286
255 also approximates the number of KSHV genomes sampled from BCBL-1.

256 **Table 1. Origin and processing results from specimens for KSHV genome analysis.**

Pt ID	Age	Sex	Plasma HIV RNA (copies/mL)	CD4 T cells/µL	Sample ID	Sample Type	% On-target		Times enriched	Fold enrichment	Mean read coverage	Mean dUMI-consensus read coverage	Genome length (excluding repeats and Ns)	#
							Pre-enrichment	Post-enrichment						v
n/a	n/a	n/a	n/a	n/a	BCBL-1	cell line	0.09%	15.61%	1x	173	7,586	286	132,676	
U003	25	M	759,635	45	**U003-C**	**Tumor**	**0.37%**	**7.15%**	**1x**	**1,925**	**15,635**	**259**	**131,292**	
					U003-o1	*Oral swab*	*<0.01%*	*35.20%*	*2x*	*2,919,559*	*47,445*	*7*	*131,102*	
					U003-o2	*Oral swab*	*0.03%*	*31.60%*	*2x*	*120,023*	*63,258*	*292*	*131,129*	
					U003-o3	*Oral swab*	*<0.01%*	*41.90%*	*2x*	*1,007,251*	*58,467*	*41*	*131,143*	
U004	37	M	277,655	85	**U004-C**	**Tumor**	**0.16%**	**17.20%**	**1x**	**10,567**	**18,787**	**32**	**131,277**	
					U004-D	**Tumor**	**0.17%**	**75.10%**	**1x**	**45,228**	**58,120**	**1,263**	**131,237**	
					U004-o1	*Oral swab*	*0.01%*	*18.90%*	*2x*	*335,500*	*36,466*	*55*	*131,277*	
U007	26	M	91,096	136	**U007-B**	**Tumor**	**0.06%**	**86.90%**	**2x**	**137,642**	**67,373**	**1,040**	**131,352**	
					U007-o1	*Oral swab*	*0.01%*	*27.60%*	*2x*	*402,068*	*41,922*	*198*	*131,126*	
U008	56	M	860,937	488	**U008-B**	**Tumor**	**0.49%**	**16.97%**	**1x**	**3,470**	**19,553**	**644**	**131,142**	
					U008-D	**Tumor**	**0.40%**	**29.98%**	**2x**	**7,534**	**25,341**	**188**	**131,102**	
					U008-o1	*Oral swab*	*<0.01%*	*23.50%*	*2x*	*846,462*	*46,611*	*111*	*131,116*	
U020	27	M	118,191	370	**U020-B**	**Tumor**	**0.70%**	**1.47%**	**1x**	**210**	**12,774**	**154**	**131,102**	
					U020-C	**Tumor**	**0.03%**	**34.38%**	**2x**	**98,720**	**15,438**	**0**	**131,471**	
					U020-o1	*Oral swab*	*<0.01%*	*77.30%*	*2x*	*3,090,161*	*7,807*	*80*	*131,115*	
					U020-o2	*Oral swab*	*<0.01%*	*51.40%*	*2x*	*3,702,047*	*6*	*0*	*128,518*	
U023	33	F	338,285	191	*U023-o1*	Oral	0.01%	*2.70%*	2x	*25,189*	19,877	1	131,122	
U030	40	M	100,184	70	**U030-C**	**Tumor**	**1.35%**	**15.68%**	**1x**	**1,162**	**36,741**	**468**	**131,282**	
					U030-o1	*Oral swab*	*<0.01%*	*5.40%*	2x	*196*	*9,402*	*0*	*128,676*	
U032	23	F	587,149	274	**U032-B**	**Tumor**	**0.03%**	**43.50%**	**2x**	**127,887**	**66,595**	**854**	**131,266**	
					U032-o1	*Oral*	*<0.01%*	*7.70%*	2x	*4,422,546*	*7,303*	*1*	*130,842*	
U034	47	F	130,375	237	**U034-B**	**Tumor**	**0.14%**	**84.40%**	2x	**41,282**	**71,473**	**1,653**	**131,248**	
					U034-C	**Tumor**	**0.13%**	**30.70%**	2x	**22,056**	**17,074**	**126**	**131,088**	
					U034-o1	*Oral swab*	*<0.01%*	*7.30%*	2x	*3,073,915*	*7,635*	*2*	*130,754*	
					U034-o2	*Oral swab*	*<0.01%*	*6.00%*	2x	*6,411,538*	*4,087*	*1*	*130,884*	

257

Eighty-one base positions (0.06%) in the BCBL1 consensus KSHV genome had detectable variants in dUMI-consensus reads, and the average frequency of minor variants was 1.35%. No variant exceeded 14% of the total dUMI-consensus reads at any position (**Fig 1C)**. No doubling of read coverage was found within the 19-kb genomic region previously reported in the BCBL-1-derived KSHV recombinant clone BAC-36 [77].

The consensus, *de novo*-assembled KSHV genome in BCBL-1 had 3 differences from the published BAC-36 sequence: a C→A change in the noncoding sequence before ORF K5 (BAC-36 position 24,630), 2 additional Gs in a homopolymer run at BAC-36 position 25,210), and a synonymous T→C change in the K7 gene (BAC-36 position 28,409). No variant bases were found in dUMI-consensus reads at the equivalent positions, indicating that the 3 BAC-36 sequence variants were not present in this passage of the BCBL-1 line at detectable levels (i.e., <1 copy per 286 genomes).

KSHV sequence derivation from tumor tissues and oral swabs

KSHV genome sequences were successfully obtained from samples provided by 9 participants with HIV-associated KS, including 12 KS tumors and 11 oral swabs. (**Table 1**). The representation of KSHV DNA in a sample was determined by ddPCR analysis of KSHV genes vIL-6, vBCL-2, RTA and LANA (**Fig 1A**) and provided as the percentage "on-target" KSHV DNA. These levels ranged from 0.03% to 1.35% (median 0.17%) in tumors, while most oral swab samples were below 0.01% on-target (**Table 1**). Following one enrichment with RNA baits, KSHV DNA corresponded to a median of 1.3% on-target, a >6,000-fold increase, and after a second enrichment a median of 24.2% on-target was obtained, for a total of 120,000-fold enrichment (**Table 1**).

Median total read coverage across KSHV genomes was 22,000 for tumors and 38,000 for oral swab samples. After collapsing mapped reads by dUMI, the median read coverage, corresponding to the number of viral genomes assessed, was 364 for tumors and 7 for oral swabs (**Table 1, S3A-B Fig**). Tumor

sample U032-B had the highest number of genomes analyzed at 1,653. We set the lowest number of reads accepted for confident assignment of variant frequencies to be 100 (**S4A Fig**)**;** below this number dUMI-consensus read coverage was judged to be too sparse. U020-C was an exception because its low mean dUMI-consensus read coverage was due to most of the KSHV genome being deleted, as discussed below. For other samples with mean dUMI-consensus read coverage below 80, all from oral swabs, the dUMI-consensus reads generated were insufficient to cover the entire KSHV genome, although whole KSHV genomes could be assembled from raw reads. Overall, read coverage was relatively uniform along the KSHV genome for most tumors (**S3A Fig**) and all adequately sampled oral swabs (**S3B Fig**).

Very few point mutations were found in dUMI-consensus reads from either tumors (**Fig 3A**) or oral swabs (**Fig 3B**). Excluding the major repeat regions, the number of positions with a detectable intrasample variant base ranged from 2 – 218 (<0.01 – 0.17%) (**Table 1**). These frequencies were lower or comparable to those in the BCBL-1 cell line (**Table 1**). The sample-consensus genome was generally the only KSHV sequence present in each sample, hence, there was no evidence for the existence of quasispecies [78]. However, in contrast to that observed in BCBL-1 viral genomes, clinical samples had detectable variation in long homopolymer runs.

Artifacts resulting from the end-repair step in DNA library preparation, which precedes the application of dUMI tags, cannot not be corrected by duplex sequencing [55,60,79]. Hence, 9 bases were removed from ends of dUMI-consensus reads before analyses, and this substantially reduced the variation observed in the raw data (data not shown). The minor base variants remaining in all samples revealed a preponderance of C→A and G→T substitutions (**S4B Fig**) as well as differences in homopolymer run lengths (**Fig 3A & B**). Most minor variants were supported by only one dUMI-consensus read. Overall, there was an inverse relationship between mean variant frequency and mean dUMI-consensus read coverage (**S4A Fig**). Thus, true minor variant frequencies could be even lower than reported here.

304

305 **KSHV genomes were virtually identical at the point mutational level between tumors and oral swab**
306 **samples from the same individual**

307 Intra-individual single nucleotide differences between tumor and oral swab samples ranged in
308 number from 0 – 2 across the entire ~131-kb genomes, not counting the major repeat regions. Notably,
309 there were almost no intra-individual polymorphisms in the KSHV hypervariable gene K1 (**Fig 3A-B**). Hence,
310 no evidence for minor KSHV variants or multi-strain infections was found in these individuals.

311 KSHV genomes were distinct across the 9 participants, with sequence differences ranging from 3.06-
312 4.85%. These genomes corresponded to K1 subtypes A5, B1 and C3 (**S5A Fig**) and K15 alleles P and M (**S5B**
313 **Fig**). While K1 and K15 are the most variable KSHV genes, polymorphisms along the rest of the genome
314 have been reported to contribute more in aggregate to the total diversity of KSHV [33,34,36]. Consistent
315 with this, maximum-likelihood phylogenetic trees using entire KSHV genomes (**S5C Fig**) were topologically
316 distinct from those of K1 or K15. Moreover, due to numerous signatures of recombination in the
317 evolutionary history of KSHV [36,48], differing phylogenies across sections of the KSHV genome may be
318 better represented by a phylogenetic network (**S5D Fig)**, in which higher degrees of conflict result in a
319 more web-like structure rather than a tree.

320

321 **Aberrant KSHV genome structures in tumors**

322 Among the 12 tumor-derived KSHV genomes examined, 7 had anomalous read coverage that shifted
323 abruptly once or twice along the viral genome (**S3A Fig**). In contrast, oral swab KSHV genomes from the
324 same individuals had uniform read coverage. This argues against preferential target capture by RNA baits
325 or other biases. Enrichment and sequencing in some were repeated, and the distinctive read coverages
326 were reproduced. Split reads accumulated at the points of abrupt shifts in read coverage and remained

after collapsing all reads by their dUMI consensus, which removes PCR chimera artifacts. Individual anomalies observed are detailed below, along with any additional evidence showing that these represented real structural aberrations in viral genomes:

Tumor 003-C. Read coverage in U003-C was high (average of 15,635) and uniform across the KSHV genome except for a 6-bp gap within the K8.1 gene intron up to the first base of the second K8.1 exon (**Fig 4A**). No read indicated a deletion, nor was any read found with its mate pair located across the 6-bp gap. This region was PCR amplified from unsheared U003-C tumor DNA using conserved primers flanking the gap (**Fig 4B**), and no PCR product was detectable. In contrast, an intact K8.1 intron sequence was amplified and sequenced from the oral swab of the same participant (**Fig 4C**).

De novo assembly revealed that the reverse complement of TR sequences continued from the deleted K8.1 intron sequences of U003-C (**Fig 4B**). The K8.1-TR junctions were confirmed by PCR with primers flanking the junctions (**Fig 4D**) and Sanger sequencing. Inversion of the 60-kb 3' half of the U003-C genome, starting inside K8.1, is a parsimonious explanation for the breakpoints.

Tumor U004-D. The first 3kb, from K1 to the end of gene ORF4, had 1.5X read coverage compared to the rest of the KSHV genome (**S3A Fig**). However, no split reads or chimeric read pairs were found to explain this result from a genome rearrangement or deletion.

Tumor U008-B and D. U008-B had 1.7X greater read coverage over a 14.8-kb segment from inside K3 to inside ORF19 (GK18 reference positions 19,168 to 33,980, **Fig 5A**), including IR1 (masked). This was corroborated by ddPCR quantitation of vBCL-2, inside the 1.7X coverage region, with 1.7 – 1.9-fold higher gene copy number in the tumor compared to vIL-6, RTA and LANA (**Table 2**).

348 **Table 2. Gene copy numbers in tumor DNA.**

Sample	vIL-6	vBCL-2	RTA	LANA
U003-C	N/A	N/A	N/A	9,664
U003-o1	127	100	120	135
U003-o2	1,298	1,238	1,452	1,628
U003-o3	191	184	169	199
U004-C	1,243	1,274	1,205	998
U004-D	4,433	4,466	4,543	4,290
U004-o1	117	128	120	112
U007-B	4,433	4,466	4,543	4,290
U007-o1	376	356	322	344
U008-B	19,140	33,629	19,910	19,195
U008-D	24,360	34,755	12,737	17,189
U008-o1	129	136	119	138
U020-B	49,600	55,850	4,550	5,500
U020-C	3,658	5,033	145	139
U020-o1	254	231	234	248
U020-o2	100	183	123	225
U023-o1	62	57	50	66
U030-C	N/A	N/A	N/A	59,730
U030-o1	2	5	5	7
U032-B	476	520	494	466
U032-o1	9	0	2	0
U034-B	1,920	1,887	1,870	1,793
U034-C	13,083	13,335	8,148	6,552
U034-o1	9	6	9	7
U034-o2	6	11	0	1

349 N/A, did not quantify

350 Inferring from split reads, the 14.8-kb segment was translocated to inside IR2 (to GK18 position
351 119,496, **Fig 5D**). This was confirmed in the unsheared tumor DNA extract by PCR and Sanger sequencing
352 using primers spanning the breakpoint (**Fig 5D & E**, lanes 4 & 5). Other primer pair combinations were
353 tested to see if there were DNA species with the 14.8-kb segment inverted, deleted in place, duplicated
354 in tandem or rearranged in other ways. None generate detectable PCR products except for primer pairs

355 showing that the 14.8-kb segment also exists in the native configuration (**Fig 5E**). Thus, the 14.8-kb
356 segment was copied into IR2 but had not been deleted from its original location.

357 In a parallel study of viral transcriptomes [80], abundant expression of a chimeric Kaposin transcript
358 fused to the 14.8-kb segment was found in tumor U008-B, consistent with the viral genome structure we
359 observed. Another tumor from the same participant, **U008-D** (**Fig 5B**), had 100% nucleotide identity and
360 was confirmed to have the same duplication and breakpoint junctions (**Fig 5F**).

361 **Tumor U020-B.** Read coverage abruptly dropped 12.8-fold over the last ~90 kb of the KSHV genome in
362 this tumor (**Fig 6A**). This was consistent with ddPCR quantitation, with vIL-6 and vBCL-2 gene amplicons
363 having 9.0 – 12.3-fold higher levels than RTA and LANA (**Table 2**). The coverage shifted before the end of
364 ORF25 (GK18 position 46,615) and reads at this breakpoint continue into TR sequences ~90 kb
365 downstream (**Fig 6C**). Thus, U020-B appeared to have KSHV genome variants with a ~90-kb deletion, or
366 formally, a 12.8X amplification of a 46-kb subgenomic region. No U020-B tumor DNA remained to allow
367 confirmation of this breakpoint.

368 **Tumor U020-C.** This tumor from participant U020 had (30-fold) shift in read coverage and different
369 breakpoints inside ORF11 and ORF18 (**Fig 6B**). ddPCR quantitation demonstrated gene copy numbers of
370 vIL-6 and vBCL-2 amplicons to be 25 – 36-fold higher than for RTA and LANA (**Table 2**). The spike in read
371 coverage occurred over a 16.2 kb region (GK18 positions 16,942 to 33,011). Again, chimeric reads were
372 found at either ends of this region continuing into TR sequences (**Fig 6D**), indicating fusion with the TR
373 (**Fig 6E**). Junction fragment-specific PCR and Sanger sequencing confirmed the 3' junction (**Fig 6G**, lane 5).
374 No PCR product was produced from the other putative breakpoint junction, TR-ORF11 (**Fig 6F**, lane 4
375 primers in K2 and TR). However, the latter result is likely due to GC-rich TR sequences being largely
376 unsuitable for primer binding. Many potential forward TR primers paired with a functional reverse TR
377 primer (**Fig 6F**, lane 5) yielded no PCR product when a control BCBL-1 DNA was used as template.

Tumor U030-C. A uniform >30,000 reads/position was observed throughout most of the KSHV genome. However, coverage dropped or was missing within the K15 gene (**S3A Fig**). The remaining K15 sequences corresponded to the K15 M-allele, which is less common than the P allele but was included in our RNA bait design (GenBank U75698). PCR amplification and Sanger sequencing of this region showed that the U030-C tumor did contain some copies of the entire M-allele K15 sequence. The U030-C sample-consensus genome was finished with this sequence, but no reads mapped to the gaps in K15. In the parallel RNAseq study of the same participants, transcripts of K15 were lacking in U030-B and C, despite being produced in all other tumor samples [80].

The same aberrant KSHV genomes are found in multiple lesions from the same individual

When breakpoint junction sequences marking an aberrant KSHV genome were confirmed by PCR in a tumor, PCR primers across those breakpoints were used to screen for the same structures in other available tumors from the same individual. In the case of U008-B and U008-D, full-length genome sequencing showed that they had the same 14.8 kb subgenomic sequence duplicated in IR2 (**Fig 5F**). These two tumors were biopsied from distinct lesions on the left leg (**S6 Fig; S2 Table**)**.** Nested PCR screening for this breakpoint junction sequence in 6 other distinct lesions (**S2 Table**) from this individual showed that no other tumors had this duplication (not shown).

In contrast, four additional tumors tested from participant U003 had the same inversion breakpoints as tumor U003-C (**Fig 7**). Moreover, no intact K8.1 sequences were detected in 2 of these 4 tumors by nested PCR of the region spanning the K8.1 intron gap (**Fig 7**). These biopsies came from distinct lesions in the left leg (**S2 Table**). Lastly, in participant U020, the ORF18-TR junction sequences found spanning the U020-C genomic deletion was not detected in the 2 other tumors tested.

401 **Mutations in sample-consensus KSHV genomes from tumors impacted protein coding sequences**

402 Among the 7 participants with KSHV sequences from at least one oral swab and one tumor, sample-
403 consensus KSHV genomes were identical in the oral and tumor samples of 2 participants and differed in 4
404 others. In the remaining participant, U004, the sample-consensus KSHV genome in one tumor was
405 identical to that in oral but the second tumor had mutations. The mutations unique to tumors were
406 typically nonsynonymous point mutations resulting in highly dissimilar residues or other mutations likely
407 to disrupt their expression (**Table 3**).

Table 3. Unique KSHV mutations observed in tumors compared to oral swabs from the same individual

Sample ID	Tumor-specific differences
U003-C	K12 synonymous mutation within miR-K10 genomic inversion starting at K8.1
U004-C	NONE
U004-D	ORF32 nonsynonymous mutation R56Q K15 nonsynonymous mutation A290P 28-nt deletion within the K8.1 promoter 3-kb segment duplication from before K1 to after ORF4
U007-B	NONE
U008-B	duplication of 14.8 kb segment around IR1 into IR2 - breakpoints inside K3 & ORF19 - genes duplicated: ORF70, K4.1, K4.2, K5, K6, K7, ORF16, ORF17, ORF17.5, ORF18
U008-D	same as U008-B
U020-B	ORF25 nonsynonymous mutation Q594K genomic deletion connecting end of ORF25 coding sequence to TR sequences, 47 kb remaining
U020-C	ORF11 nonsynonymous mutation T396P K3 nonsynonymous mutation F88L in transmembrane domain K8.1 nonsense mutation at start of 2nd exon genomic deletion leaving only 16 kb segment surrounding IR1 connected to TR sequences - breakpoints inside ORF11 and ORF18 - ~30X coverage for: K2, ORF2, K3, ORF70, K4, K4.1, K4.2, K5, K6, K7, ORF16, ORF7, ORF17.5
U032-B	ORF63 nonsynonymous mutation T848A
U034-B	NONE
U034-C	NONE

Several tumor-unique mutations or genome aberrations occurred in structural genes (**S3 Table**), and frequently involved the K8.1 gene, which encodes an envelope glycoprotein. The U003 inversion breakpoint cleaved the K8.1 gene. U004-D had an R56Q mutation in its ORF32 tegument protein coding sequence, as well as a 28-nt deletion in the promoter region of K8.1 (**S7A Fig**). The deletion was after the K8.1 core promoter sequence [81], but encompassed the K8.1 transcription start site [82]. The ORF25 major capsid protein in U020-B had a Q594K mutation, in addition to the U020-B genomic deletion that started downstream of ORF25. U020-C had a nonsense mutation at the beginning of the second K8.1 exon. Finally, U032-B had a T848A mutation in ORF63, a tegument protein.

The only intra-host synonymous point mutation observed was in ORF K12 of U003-C (GK18 position 118,082). This C to T change occurred within the oncogenic microRNA K10 (miR-K10) sequence in the Kaposin A transcript. The three oral swab samples from this participant maintained the consensus C at this position (**S7B Fig**), whereas the 4 other tumors from this participant examined had T at this position, with tumor U003-G having a minor population of viruses with the consensus C (**S7C Fig**). Among published KSHV genomes, only ZM106 (GenBank KT271458), also derived from a KS tumor, had a T at this position.

Lack of evidence for integration of KSHV sequences into human chromosomes

No *de novo*-assembled scaffolds, split reads or improperly-paired read mappings suggested any instance of KSHV sequences fused to human DNA. Nevertheless, attempts were made to systematically search for human-KSHV chimeric sequences. The methods employed were those used to screen for all integrated herpesviruses sequences in public databases [83] and EBV integration sites in in primary gastric and nasopharyngeal carcinomas [84]. The KSHV genome inversions, duplications and deletions described above were detected by LUMPY with high confidence values. In contrast, putative breakpoints that joined

human and KSHV sequences were supported by only tens of reads, about 2 orders of magnitude lower in number, and often involved LANA repeats into low-complexity human repeat sequences.

Co-infection with EBV detected predominantly in oral swabs

Some scaffolds during *de novo* assembly correspond to EBV sequences. Nearly all oral swab samples yielded multiple EBV-mapping scaffolds up to 73 kb, with no region of the EBV genome over-represented. In contrast, EBV-sequences were detected in only 5 of 12 tumors, and in all cases were sequences flanking the EBNA-1 repeat (**S8 Fig**). The proportion of reads mapping to EBV in oral swabs ranged from 0 - 33%, median 1.8%, whereas in tumors the range was from 0 - 0.5%, median 0.002% (**S4 Table**). No other eukaryote viruses were identified, including HIV, with which every participant was known to be infected.

Discussion

This study is the first to explore KSHV intra-host diversity at the whole genome level and provided an unprecedented level of precision to herpesvirus genome sequence analysis in clinical specimens, with >100 essentially error-free genome sequences obtained from most tumors. KSHV genomes were obtained from 11 oral swabs and 12 KS tumors of 9 Ugandan adults with KS. By incorporating dUMI, PCR misincorporation errors and template switching artifacts were substantially eliminated, permitting detection variants as infrequent as 0.01% and a theoretical error rate of $1/10^{-9}$, or approximately the DNA replication error rate in eukaryotic cells [55].

There were no signs of KSHV quasispecies, consistent with large dsDNA viruses having the lowest mutation rates among viruses [85]. Less than 0.01% of all base positions in the 131-kb KSHV genomes (excluding the major repeat regions) were found to have a detectable variant, typically supported by only

452 one dUMI-consensus read. This exceedingly low intra-sample variation is within the published resolution
453 of duplex sequencing [55]. While there are reports of intra-host KSHV variability in certain KSHV-endemic
454 populations [38], in children [43], in iatrogenic settings [39–41] and in blood of AIDS-KS patients [42],
455 these findings were arrived at by Sanger sequencing of PCR amplicon clones of hypervariable regions in
456 K1 or other genes. Such protocols are more likely to detect errors that occurred during PCR. Our study
457 found virtually no intra-sample or intra-host diversity even at K1 in the 9 individuals examined.
458 Recombination is evident in the evolutionary history of KSHV [36,44,48,49]; hence, co-infections by
459 multiple KSHV strains must occur, if sporadically.

460 The most striking observation in this study was the frequency and tumor-specificity of aberrant KSHV
461 genomes, summarized in **Figure 8**. Up to 7 of the 12 KS tumors examined had major inversions, deletions
462 or duplications comprising the majority of KSHV genomes in those tumors. In stark contrast, no aberrant
463 genome structures were found in oral swabs. It is unclear whether the tumor-specific mutations observed
464 were required for tumorigenesis or tumor persistence, or whether they were a consequence of localized
465 genomic instability known to occur in tumors [86], but several observations suggest that these changes
466 were not random. Rearrangement breakpoints and other mutations almost always occurred inside coding
467 sequences of lytic genes (**Tables 3 & S3**), with many truncating protein coding sequences. These mutations
468 may have been selected for, if for instance expressing these proteins exposed host cells to immune
469 targeting. Mutations that disrupt genes may also indicate that those genes were not necessary for
470 sustaining tumorigenic growth. Conversely, regions of the KSHV genome that were duplicated,
471 conspicuously intact or translocated next to strong promoters (as in 008-B and 008-D) may point to KSHV
472 genes that are influential in driving tumor cell proliferation.

473 The genomic region around IR1 featured prominently in genomic rearrangements in 4 tumors,
474 potentially leading to their over-expression relative to other KSHV genes. For example, tumors U008-B
475 and U008-D had a 14.8-kb portion of their genomes, from inside K3 to ORF19, duplicated into within IR2

476 (**Fig 5**). In a parallel RNAseq study, tumor U008-B had been found to abundantly express a chimeric
477 transcript of the 14.8 kb section fused to IR2 sequences transcribed from a strong latency-associated
478 promoter [80]. Distinct deletions were observed in tumors U020-B and U020-C from another participant,
479 but the genomic regions retained, aside from TR sequences, again included the IR1 region (**Fig 6**). There is
480 evidence to suggest that KSHV lytic gene expression is crucial to KS pathogenesis [87], and that residual
481 lytic gene expression plays a role in latent KSHV persistence *in vivo* [88]. IR1 is one of the origins of lytic
482 replication, and transcripts around IR1 are among the most highly expressed in KS tumors [80]. These
483 include two long non-coding RNAs that have indispensable roles during lytic reactivation of KSHV, T1.4
484 [67,89,90] and PAN [91–93]. PAN has been shown to interact with promoters of cellular genes involved in
485 inflammation, cell cycle regulation and metabolism, and exogenous expression of PAN alone enhanced
486 cell growth phenotypes [94]. Recently, virally-encoded circular RNAs encoded within PAN were discovered
487 to be abundant in clinical samples and were inducible in KSHV-infected cell lines [95–97]. Other non-
488 coding transcripts are potentially expressed from this region but their biological significance is unknown
489 [93,98]. Finally, most ORFs encoded in the 14.8-kb retained region are intermediate-early or delayed-early
490 lytic genes that may have functions in subverting adaptive (K5/MIR2) [99,100] or innate immunity
491 (K4/vCCL-2, K4.1/vCCL-3, K4.2 and K6/vCCL-1)[100–103], and apoptosis (K7 and ORF16/vBCL-2) [100,103].

492 Other than genomic rearrangements, sample-consensus KSHV genomes in tumors and oral swabs
493 within the same individuals differed by at most two point mutations or a short deletion. No mutations
494 occurred in intergenic regions, and almost all were nonsynonymous changes resulting in highly dissimilar
495 amino acids. Notably, the sole intra-host synonymous mutation found occurred inside K12/Kaposin A of
496 tumor U003-C (GK18 position 118,082), within the embedded microRNA miR-K10. The oral swab
497 counterpart maintained the database consensus. Expression of the Kaposin A transcript is tumorigenic
498 [104] and a single base change has been observed to abolish this effect [105], although a different
499 mutation was observed here.

The late lytic gene K8.1 was found to be mutated in KS tumors from 3 individuals. U003-C had an inversion breakpoint at the K8.1 intron, U004-D had a 28-bp deletion ending at 4 bases upstream of the first K8.1 exon, and U020-C had a nonsense mutation at the start of the second K8.1 exon (**Fig 8**). Furthermore, intact K8.1 gene sequences were undetectable by PCR in most of participant 003's tumors tested (**Figs 4C & 7**). Truncations in K8.1 had been reported previously, all from KS tumor isolates. The original GK18 isolate had a 74-bp deletion at the 3' end (see GenBank ID AF148805 K8.1 annotation); the Zambian isolate ZM124 (GenBank ID: KT271466) had a 25-nt deletion resulting in a frameshift and premature stop [33]; finally, Japanese isolate Miyako1 has a stop codon early in its first exon (see GenBank ID LC200586 miscellaneous annotation). Collectively, these findings suggest selection pressure against K8.1 expression in tumors.

Gene K8.1 encodes an envelope glycoprotein that interacts with heparin sulfate for attachment [106–109]. It is not required for entry into endothelial [107] or 293 cells [110], although it had recently been shown to be necessary for infection of primary and cultured B-cells [111]. The K8.1 protein is often used as an indicator of the late lytic stage and is among the most immunogenic KSHV proteins [112–114]. It is therefore conceivable that the preponderance of K8.1 mutations might be due to potent immune targeting of cells expressing K8.1 glycoproteins.

Some of the aberrant KSHV genomes we observed would be unable to produce infectious virions, yet the same viral genome rearrangements were sometimes found in multiple lesions. All 5 tumors tested of Participant U003 had the K8.1-TR junction present, but intact K8.1 sequence was detectable by PCR in only 3 (**Fig 7**). Participant U008 had the same sequence duplication in 2 distinct tumor lesions (**Figs 5 & S6)**. Thus, spread of these mutated genomes could have occurred by tumor metastasis or with a helper virus.

Detection of aberrant KSHV genomes is not without precedent. The first whole genome sequence of KSHV published reported a 33-kb portion of the KSHV unique central region duplicated into the TR region [115]. A study of 16 tumor-derived KSHV whole genomes from Zambia reported 4 that had regions with 3-fold more coverage than the sample average, although these were not examined [33] . A PCR screen for some KSHV genes showed that some KS tumors and KSHV-infected B-cell lines can harbor deleted KSHV genomes [116], and one such B-cell line proliferated faster than the parental BCBL-1 line [116]. The infecting KSHV had an 82-kb deletion from the 5' end of its genome, was lytic replication-incompetent, and could be packaged by a helper virus.

The LANA protein tethers the KSHV episome and is required for maintaining KSHV latency and for latent replication of TR sequence-containing plasmids [117,118], but the deleted 020-B and 020-C genomes had the entire latency locus, including LANA, missing. Remaining intact KSHV genomes in the same cell could be supplying LANA, as latently infected cells frequently harbor multiple KSHV copies per cell [118,119], and KSHV episomes are inherited by daughter cells in clusters [120].

Two remaining observations of note in this study were that, in contrast to larger studies of EBV in tumors [84,121], we found no integrations of KSHV into human chromosomal DNA. Secondly, our cohort was infected also with EBV, as seen in the abundance of sequence reads corresponding to EBV especially in oral swabs. Tellingly, the rare EBV sequences found in tumors were consistently of EBNA-1, whose GC-rich repeat domain has nucleotide homology to LANA repeats [122] and hence were probably co-captured by the RNA baits. No sequences of HIV or other eukaryote viruses were detected.

In summary, highly accurate deep sequencing of whole KSHV genomes in paired oral swab and KS tumors from individuals with advanced KS were virtually identical at the point mutational level. Where there were differences, the oral viruses had the database consensus genotype while tumor viruses had novel mutations. KS tumors can harbor KSHV with genomic aberrations or other mutations that may alter

lytic gene expression, and these viral mutations can be shared by distinct KS tumors within an individual. Our study points to associations with KS tumors of the region surrounding IR1 and the K8.1 gene. As inactivating mutations seem to be a frequent feature of tumor-derived KSHV, whole genome or targeted sequencing may reveal more viral genomic regions important to the pathogenesis or persistence of KS.

Acknowledgements

We thank the study participants who contributed invaluable specimens that make this study possible. We also thank D. Depledge and J. Breuer for sharing their RNA bait sequences for capturing KSHV genomic sequences and M. Lagunoff for kindly providing us with an early passage BCBL-1 cell line.

References

1. Engels EA, Pfeiffer RM, Goedert JJ, Virgo P, McNeel TS, Scoppa SM, et al. Trends in cancer risk among people with AIDS in the United States 1980-2002. AIDS. 2006;20: 1645–54. doi:10.1097/01.aids.0000238411.75324.59 PMID: 16868446

2. Mbulaiteye SM, Katabira ET, Wabinga H, Parkin DM, Virgo P, Ochai R, et al. Spectrum of cancers among HIV-infected persons in Africa: the Uganda AIDS-Cancer Registry Match Study. Int J cancer. 2006;118: 985–90. doi:10.1002/ijc.21443 PMID: 16106415

3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61: 69–90. doi:10.3322/caac.20107 PMID: 21296855

4. Sarid R, Gao S-J. Viruses and human cancer: from detection to causality. Cancer Lett. 2011;305: 218–27. doi:10.1016/j.canlet.2010.09.011 PMID: 20971551

5. Moore PS, Chang Y. The conundrum of causality in tumor virology: the cases of KSHV and MCV. Semin Cancer Biol. 2014;26: 4–12. doi:10.1016/j.semcancer.2013.11.001 PMID: 24304907

6. Koelle DM, Huang M, Chandran B, Vieira J, Piepkorn M, Corey L. Frequent Detection of Kaposi's Sarcoma—Associated Herpesvirus (Human Herpesvirus 8) DNA in Saliva of Human Immunodeficiency Virus-Infected Men: Clinical and Immunologic Correlates. J Infect Dis. 1997. doi:10.1086/514045 PMID: 9207354

7. Casper C, Lampinen TM, Morrow RA, Corey L, Wald A. HIV Infection and Human Herpesvirus-8 Oral Shedding. 2004;35: 233–238.

8. Phipps W, Saracino M, Selke S, Huang M-L, Jaoko W, Mandaliya K, et al. Oral HHV-8 replication

among women in Mombasa, Kenya. J Med Virol. 2014;86: 1759–65. doi:10.1002/jmv.23941 PMID: 24692069

9. Mesri EA, Cesarman E, Boshoff C. Kaposi's sarcoma and its associated herpesvirus. Nature Reviews Cancer. NIH Public Access; 2010. pp. 707–719. doi:10.1038/nrc2888 PMID: 20865011

10. Haley CT, Mui UN, Vangipuram R, Rady PL, Tyring SK. Human Oncoviruses: Mucocutaneous Manifestations, Pathogenesis, Therapeutics, and Prevention (Part I: Papillomaviruses and Merkel cell polyomavirus). J Am Acad Dermatol. 2018. doi:10.1016/J.JAAD.2018.09.062

11. Sitki-Green D, Edwards RH, Webster-Cyriaque J, Raab-Traub N. Identification of Epstein-Barr virus strain variants in hairy leukoplakia and peripheral blood by use of a heteroduplex tracking assay. J Virol. 2002;76: 9645–56. doi:10.1128/jvi.76.19.9645-9656.2002 PMID: 12208943

12. Edwards RH, Sitki-Green D, Moore DT, Raab-Traub N. Potential selection of LMP1 variants in nasopharyngeal carcinoma. J Virol. 2004;78: 868–81. doi:10.1128/jvi.78.2.868-881.2004 PMID: 14694118

13. Weiss ER, Lamers SL, Henderson JL, Melnikov A, Somasundaran M, Garber M, et al. Early Epstein-Barr Virus Genomic Diversity and Convergence toward the B95.8 Genome in Primary Infection. J Virol. 2017;92: JVI.01466-17. doi:10.1128/jvi.01466-17 PMID: 29093087

14. Kwok H, Tong AHY, Lin CH, Lok S, Farrell PJ, Kwong DLW, et al. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. Lo K-W, editor. PLoS One. 2012;7: e36939. doi:10.1371/journal.pone.0036939 PMID: 22590638

15. Kwok H, Wu CW, Palser AL, Kellam P, Sham PC, Kwong DLW, et al. Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. J Virol. 2014;88: 10662–72. doi:10.1128/JVI.01665-14 PMID: 24991008

16. Tsai M-H, Lin X, Shumilov A, Bernhardt K, Feederle R, Poirey R, et al. The biological properties of different Epstein-Barr virus strains explain their association with various types of cancers. Oncotarget. 2017;8: 10238–10254. doi:10.18632/oncotarget.14380 PMID: 28052012

17. Dheekollu J, Malecka K, Wiedmer A, Delecluse H-J, Chiang AKS, Altieri DC, et al. Carcinoma-risk variant of EBNA1 deregulates Epstein-Barr Virus episomal latency. Oncotarget. 2017;8: 7248–7264. doi:10.18632/oncotarget.14540 PMID: 28077791

18. Church TM, Verma D, Thompson J, Swaminathan S. Efficient Translation of Epstein-Barr Virus (EBV) DNA Polymerase Contributes to the Enhanced Lytic Replication Phenotype of M81 EBV. J Virol. 2018;92. doi:10.1128/JVI.01794-17 PMID: 29263273

19. Liu Q, Han A, You S, Yang Q, Liang Y, Dong Y. The association of genomic variation of Epstein-Barr virus BamHI F fragment with the proliferation of nasopharyngeal carcinoma. APMIS. 2010;118: 657–64. doi:10.1111/j.1600-0463.2010.02642.x PMID: 20718717

20. Tsai M-H, Raykova A, Klinke O, Bernhardt K, Gärtner K, Leung CS, et al. Spontaneous lytic replication and epitheliotropism define an Epstein-Barr virus strain found in carcinomas. Cell Rep. 2013;5: 458–70. doi:10.1016/j.celrep.2013.09.012 PMID: 24120866

21. Rezaee SAR, Cunningham C, Davison AJ, Blackbourn DJ. Kaposi's sarcoma-associated herpesvirus immune modulation: an overview. J Gen Virol. 2006;87: 1781–804. doi:10.1099/vir.0.81919-0 PMID: 16760382

22. Direkze S, Laman H. Regulation of growth signalling and cell cycle by Kaposi's sarcoma-associated

herpesvirus genes. Int J Exp Pathol. 2004;85: 305–19. doi:10.1111/j.0959-9673.2004.00407.x PMID: 15566428

23. Coscoy L. Immune evasion by Kaposi's sarcoma-associated herpesvirus. Nat Rev Immunol. 2007;7: 391–401. doi:10.1038/nri2076 PMID: 17457345

24. Ganem D. KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. J Clin Invest. 2010;120: 939–49. doi:10.1172/JCI40567 PMID: 20364091

25. Cordiali-Fei P, Trento E, Giovanetti M, Lo Presti A, Latini A, Giuliani M, et al. Analysis of the ORFK1 hypervariable regions reveal distinct HHV-8 clustering in Kaposi's sarcoma and non-Kaposi's cases. J Exp Clin Cancer Res. 2015;34: 1. doi:10.1186/s13046-014-0119-0 PMID: 25592960

26. Boralevi F, Masquelier B, Denayrolles M, Dupon M, Pellegrin JL, Ragnaud JM, et al. Study of human herpesvirus 8 (HHV-8) variants from Kaposi's sarcoma in France: is HHV-8 subtype A responsible for more agressive tumors? J Infect Dis. 1998;178: 1546–7. doi:10.1086/314471 PMID: 9780286

27. Mancuso R, Biffi R, Valli M, Bellinvia M, Athanasia T, Ferrucci S, et al. HHV8 a subtype is associated with rapidly evolving classic Kaposi's sarcoma. J Med Virol. 2008;80: 2153–2160. doi:10.1002/jmv.21322 PMID: 19040293

28. Tozetto-Mendoza TR, Ibrahim KY, Tateno AF, Oliveira CM de, Sumita LM, Sanchez MCA, et al. Genotypic distribution of HHV-8 in AIDS individuals without and with Kaposi sarcoma: Is genotype B associated with better prognosis of AIDS-KS? Medicine (Baltimore). 2016;95: e5291. doi:10.1097/MD.0000000000005291 PMID: 27902590

29. Cook PM, Whitby D, Calabro ML, Luppi M, Kakoola DN, Hjalgrim H, et al. Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. AIDS. 1999;13: 1165–1176. doi:10.1097/00002030-199907090-00004 PMID: 10416519

30. Lacoste V, Judde JG, Brière J, Tulliez M, Garin B, Kassa-Kelembho E, et al. Molecular epidemiology of human herpesvirus 8 in Africa: Both B and A5 K1 genotypes, as well as the M and P genotypes of K14.1/K15 loci, are frequent and widespread. Virology. 2000;278: 60–74. doi:10.1006/viro.2000.0629 PMID: 11112482

31. Kadyrova E, Lacoste V, Duprez R, Pozharissky K, Molochkov V, Huerre M, et al. Molecular epidemiology of Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 strains from Russian patients with classic, posttransplant, and AIDS-associated Kaposi's sarcoma. J Med Virol. 2003;71: 548–556. doi:10.1002/jmv.10530 PMID: 14556268

32. Cordiali-Fei P, Trento E, Giovanetti M, Lo Presti A, Latini A, Giuliani M, et al. Analysis of the ORFK1 hypervariable regions reveal distinct HHV-8 clustering in Kaposi's sarcoma and non-Kaposi's cases. J Exp Clin Cancer Res. 2015;34: 1. doi:10.1186/s13046-014-0119-0 PMID: 25592960

33. Olp LN, Jeanniard A, Marimo C, West JT, Wood C. Whole-Genome Sequencing of Kaposi's Sarcoma-Associated Herpesvirus from Zambian Kaposi's Sarcoma Biopsy Specimens Reveals Unique Viral Diversity. J Virol. 2015;89: 12299–12308. doi:10.1128/JVI.01712-15.Editor PMID: 26423952

34. Awazawa R, Utsumi D, Katano H, Awazawa T, Miyagi T, Hayashi K, et al. High Prevalence of Distinct Human Herpesvirus 8 Contributes to the High Incidence of Non-acquired Immune Deficiency Syndrome-Associated Kaposi's Sarcoma in Isolated Japanese Islands. J Infect Dis. 2017;216: 850–858. doi:10.1093/infdis/jix424

35. Osawa M, Mine S, Ota S, Kato K, Sekizuka T, Kuroda M, et al. Establishing and characterizing a new primary effusion lymphoma cell line harboring Kaposi's sarcoma–associated herpesvirus. Infect Agent Cancer. 2016;11: 37. doi:10.1186/s13027-016-0086-5

36. Sallah N, Palser AL, Watson SJ, Labo N, Asiki G, Marshall V, et al. Genome-Wide Sequence Analysis of Kaposi Sarcoma-Associated Herpesvirus Shows Diversification Driven by Recombination. J Infect Dis. 2018;218: 1700–1710. doi:10.1093/infdis/jiy427 PMID: 30010810

37. Caro-Vegas C, Sellers S, Host KM, Seltzer J, Landis J, Fischer WA, et al. Runaway Kaposi Sarcoma-associated herpesvirus replication correlates with systemic IL-10 levels. Virology. 2020;539: 18–25. doi:10.1016/j.virol.2019.10.002

38. Beyari MM, Hodgson TA, Cook RD, Kondowe W, Molyneux EM, Scully CM, et al. Multiple Human Herpesvirus–8 Infection. J Infect Dis. 2003;188: 678–689. doi:10.1086/377504 PMID: 12934184

39. Al-Otaibi LM, Ngui SL, Scully CM, Porter SR, Teo CG. Salivary human herpesvirus 8 shedding in renal allograft recipients with Kaposi's sarcoma. J Med Virol. 2007;79: 1357–1365. doi:10.1002/jmv.20929 PMID: 17607792

40. Al-Otaibi LM, Al-Sulaiman MH, Teo CG, Porter SR. Extensive oral shedding of human herpesvirus 8 in a renal allograft recipient. Oral Microbiol Immunol. 2009;24: 109–115. doi:10.1111/j.1399-302X.2008.00481.x PMID: 19239637

41. Al-Otaibi LM, Moles DR, Porter SR, Teo CG. Human herpesvirus 8 shedding in the mouth and blood of hemodialysis patients. J Med Virol. 2012;84: 792–797. doi:10.1002/jmv.23245 PMID: 22431028

42. Leao JC, de Faria ABS, Fonseca DDD, Gueiros LAM, Silva IHM, Porter SR. Intrahost genetic variability of human herpes virus-8. J Med Virol. 2013;85: 636–645. doi:10.1002/jmv PMID: 26928661

43. Mbulaiteye S, Marshall V, Bagni RK, Wang C, Mbisa G, Bakaki PM, et al. Molecular Evidence for Mother-to-Child Transmission of Kaposi Sarcoma–Associated Herpesvirus in Uganda and K1 Gene Evolution within the Host. J Infect Dis. 2006;193: 1250–1257. doi:10.1086/503052 PMID: 16586362

44. Zong J, Ciufo DM, Viscidi R, Alagiozoglou L, Tyring S, Rady P, et al. Genotypic analysis at multiple loci across Kaposi's sarcoma herpesvirus (KSHV) DNA molecules: Clustering patterns, novel variants and chimerism. J Clin Virol. 2002;23: 119–148. doi:10.1016/S1386-6532(01)00205-0 PMID: 11595592

45. Zong J-C, Arav-Boger R, Alcendor DJ, Hayward GS. Reflections on the interpretation of heterogeneity and strain differences based on very limited PCR sequence data from Kaposi's sarcoma-associated herpesvirus genomes. J Clin Virol. 2007;40: 1–8. doi:10.1016/j.jcv.2007.06.012 PMID: 17698410

46. Stebbing J, Wilder N, Ariad S, Abu-Shakra M. Lack of intra-patient strain variability during infection with Kaposi's sarcoma-associated herpesvirus. Am J Hematol. 2001;68: 133–4. doi:10.1002/ajh.1165 PMID: 11559954

47. Meng YX, Spira TJ, Bhat GJ, Birch CJ, Druce JD, Edlin BR, et al. Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. Virology. 1999;261: 106–119. doi:10.1006/viro.1999.9853 PMID: 10441559

48. Hayward GS, Zong JC. Modern evolutionary history of the human KSHV genome. Curr Top

Microbiol Immunol. 2007;312: 1–42. PMID: 17089792

49. Poole LJ, Zong JC, Ciufo DM, Alcendor DJ, Cannon JS, Ambinder R, et al. Comparison of genetic variability at multiple loci across the genomes of the major subtypes of Kaposi's sarcoma-associated herpesvirus reveals evidence for recombination and for two distinct types of open reading frame K15 alleles at the right-hand end. J Virol. 1999;73: 6646–60. PMID: 10400762

50. Kakoola DN, Sheldon J, Byabazaire N, Bowden RJ, Katongole-Mbidde E, Schulz TF, et al. Recombination in human herpesvirus-8 strains from Uganda and evolution of the K15 gene. J Gen Virol. 2001;82: 2393–2404. doi:10.1099/0022-1317-82-10-2393 PMID: 11562533

51. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013;41: e67. doi:10.1093/nar/gks1443 PMID: 23303777

52. Chen G, Mosier S, Gocke CD, Lin M-T, Eshleman JR. Cytosine Deamination Is a Major Cause of Baseline Noise in Next-Generation Sequencing. Mol Diagn Ther. 2014;18: 587–593. doi:10.1007/s40291-014-0115-2 PMID: 25091469

53. Ma X, Shao Y, Tian L, Flasch DA, Mulder HL, Edmonson MN, et al. Analysis of error profiles in deep next-generation sequencing data. Genome Biol. 2019;20: 50. doi:10.1186/s13059-019-1659-6

54. Park G, Park JK, Shin S-H, Jeon H-J, Kim NKD, Kim YJ, et al. Characterization of background noise in capture-based targeted sequencing data. Genome Biol. 2017;18: 136. doi:10.1186/s13059-017-1275-2 PMID: 28732520

55. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A. 2012;109: 14508–13. doi:10.1073/pnas.1208715109 PMID: 22853953

56. Krown SE, Metroka C, Wernz JC. Kaposi's sarcoma in the acquired immune deficiency syndrome: a proposal for uniform evaluation, response, and staging criteria. AIDS Clinical Trials Group Oncology Committee. J Clin Oncol. 1989;7: 1201–1207. PMID: 2671281

57. Johnston C, Orem J, Okuku F, Kalinaki M, Saracino M, Katongole-Mbidde E, et al. Impact of HIV infection and Kaposi sarcoma on human herpesvirus-8 mucosal replication and dissemination in Uganda. PLoS One. 2009;4: e4222. doi:10.1371/journal.pone.0004222 PMID: 19156206

58. Ryncarz AJ, Goddard J, Wald A, Huang ML, Roizman B, Corey L. Development of a high-throughput quantitative assay for detecting herpes simplex virus DNA in clinical samples. J Clin Microbiol. 1999;37: 1941–7. PMID: 10325351

59. Zaniello B, Huang M-L, Cheng A, Selke S, Wald A, Jerome KR, et al. Consistent viral DNA quantification after prolonged storage at ambient temperature. J Virol Methods. 2016;228: 91–94. doi:10.1016/j.jviromet.2015.11.010

60. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc. 2014;9: 2586–2606. doi:10.1038/nprot.2014.170 PMID: 25299156

61. Depledge DP, Palser AL, Watson SJ, Lai IY-C, Gray ER, Grant P, et al. Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. PLoS One. 2011;6. doi:10.1371/JOURNAL.PONE.0027805 PMID: 22125625

62. Thompson JR, Marcelino LA, Polz MF. Heteroduplexes in mixed-template amplifications:

formation, consequence and elimination by "reconditioning PCR". Nucleic Acids Res. 2002;30: 2083–8. doi:10.1093/nar/30.9.2083 PMID: 11972349

63. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. 2011; 2011.

64. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.

65. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19: 455–77. doi:10.1089/cmb.2012.0021 PMID: 22506599

66. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012;13: R56. doi:10.1186/gb-2012-13-6-r56 PMID: 22731987

67. Wang Y, Li H, Chan MY, Zhu FX, Lukac DM, Yuan Y. Kaposi's Sarcoma-Associated Herpesvirus ori-Lyt-Dependent DNA Replication: cis-Acting Requirements for Replication and ori-Lyt-Associated RNA Transcription. J Virol. 2004;78: 8615–8629. doi:10.1128/JVI.78.16.8615-8629.2004 PMID: 15280471

68. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–80. doi:10.1093/molbev/mst010 PMID: 23329690

69. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–3. doi:10.1093/bioinformatics/btu033 PMID: 24451623

70. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution. Oxford University Press; 2006. pp. 254–267. doi:10.1093/molbev/msj030 PMID: 16221896

71. Katz JP, Pipas JM. SummonChimera infers integrated viral genomes with nucleotide precision from NGS data. BMC Bioinformatics. 2014;15: 348. doi:10.1186/s12859-014-0348-4 PMID: 25331652

72. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: Ultra-fast personal genome analysis and interpretation. Nat Methods. 2015;12: 966–968. doi:10.1038/nmeth.3505 PMID: 26258291

73. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15: R84. doi:10.1186/gb-2014-15-6-r84 PMID: 24970577

74. Zhou F-C, Zhang Y-J, Deng J-H, Wang X-P, Pan H-Y, Hettler E, et al. Efficient infection by a recombinant Kaposi's sarcoma-associated herpesvirus cloned in a bacterial artificial chromosome: application for genetic analysis. J Virol. 2002;76: 6185–96. doi:10.1128/JVI.76.12.6185-6196.2002 PMID: 12021352

75. Renne R, Zhong W, Herndier B, McGrath M, Abbey N, Kedes D, et al. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. Nat Med. 1996;2: 342–346. doi:10.1038/nm0396-342

76. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2012;9: 72–74. doi:10.1038/nmeth.1778 PMID: 22101854

77. Yakushko Y, Hackmann C, Günther T, Rückert J, Henke M, Koste L, et al. Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome contains a duplication of a long unique-region fragment within the terminal repeat region. J Virol. 2011;85: 4612–7. doi:10.1128/JVI.02068-10 PMID: 21307197

78. Eigen M. On the nature of virus quasispecies. Trends Microbiol. 1996;4: 216–218. doi:10.2105/AJPH.27.11.1160 PMID: 8795155

79. Peng Q, Xu C, Kim D, Lewis M, DiCarlo J, Wang Y. Targeted Single Primer Enrichment Sequencing with Single End Duplex-UMI. Sci Rep. 2019;9: 4810. doi:10.1038/s41598-019-41215-z PMID: 30886209

80. Rose TM, Bruce AG, Barcy S, Fitzgibbon M, Matsumoto LR, Ikoma M, et al. Quantitative RNAseq analysis of Ugandan KS tumors reveals KSHV gene expression dominated by transcription from the LTd downstream latency promoter. Schulz TF, editor. PLoS Pathog. 2018;14: e1007441. doi:10.1371/journal.ppat.1007441 PMID: 30557332

81. Tang S, Yamanegi K, Zheng Z-M. Requirement of a 12-base-pair TATT-containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-associated herpesvirus K8.1 late promoter. J Virol. 2004;78: 2609–14. doi:10.1128/jvi.78.5.2609-2614.2004 PMID: 14963167

82. Tang S, Zheng Z-M. Kaposi's sarcoma-associated herpesvirus K8 exon 3 contains three 5'-splice sites and harbors a K8.1 transcription start site. J Biol Chem. 2002;277: 14547–56. doi:10.1074/jbc.M111308200 PMID: 11832484

83. Cantalupo PG, Katz JP, Pipas JM. Viral sequences in human cancer. Virology. 2018;513: 208–216. doi:10.1016/j.virol.2017.10.017 PMID: 29107929

84. Xu M, Zhang W-L, Zhu Q, Zhang S, Yao Y-Y, Xiang T, et al. Genome-wide profiling of Epstein-Barr virus integration by targeted sequencing in Epstein-Barr virus associated malignancies. Theranostics. 2019;9: 1115–1124. doi:10.7150/thno.29622 PMID: 30867819

85. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci. 2016;73: 4433–4448. doi:10.1007/s00018-016-2299-6 PMID: 27392606

86. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144: 27–40. doi:10.1016/j.cell.2010.11.055 PMID: 21215367

87. Andrei G, Snoeck R. Kaposi's sarcoma-associated herpesvirus: The role of lytic replication in targeted therapy. Current Opinion in Infectious Diseases. 2015. pp. 611–624. doi:10.1097/QCO.0000000000000213 PMID: 26524334

88. Grundhoff A, Ganem D. Inefficient establishment of KSHV latency suggests an additional role for continued lytic replication in Kaposi sarcoma pathogenesis. J Clin Invest. 2004;113: 124–36. doi:10.1172/JCI17803 PMID: 14702116

89. Wang Y, Tang Q, Maul GG, Yuan Y. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator. J Virol. 2006;80: 12171–86. doi:10.1128/JVI.00990-06 PMID: 17020951

90. Liu D, Wang Y, Yuan Y. Kaposi's Sarcoma-Associated Herpesvirus K8 Is an RNA Binding Protein That Regulates Viral DNA Replication in Coordination with a Noncoding RNA. J Virol. 2018;92. doi:10.1128/JVI.02177-17 PMID: 29321307

91. Campbell M, Kung H-J, Izumiya Y. Long non-coding RNA and epigenetic gene regulation of KSHV.

826 Viruses. 2014;6: 4165–77. doi:10.3390/v6114165 PMID: 25375882

827 92. Conrad NK. New insights into the expression and functions of the Kaposi's sarcoma-associated
828 herpesvirus long noncoding PAN RNA. Virus Res. 2016;212: 53–63.
829 doi:10.1016/j.virusres.2015.06.012 PMID: 26103097

830 93. Chavez-Calvillo G, Martin S, Hamm C, Sztuba-Solinska J. The Structure-To-Function Relationships
831 of Gammaherpesvirus-Encoded Long Non-Coding RNAs and Their Contributions to Viral
832 Pathogenesis. Non-coding RNA. 2018;4. doi:10.3390/ncrna4040024 PMID: 30261651

833 94. Rossetto CC, Tarrant-Elorza M, Verma S, Purushothaman P, Pari GS. Regulation of Viral and
834 Cellular Gene Expression by Kaposi's Sarcoma-Associated Herpesvirus Polyadenylated Nuclear
835 RNA. J Virol. 2013;87: 5540–5553. doi:10.1128/jvi.03111-12

836 95. Tagawa T, Gao S, Koparde VN, Gonzalez M, Spouge JL, Serquiña AP, et al. Discovery of Kaposi's
837 sarcoma herpesvirus-encoded circular RNAs and a human antiviral circular RNA. Proc Natl Acad
838 Sci U S A. 2018;115: 12805–12810. doi:10.1073/pnas.1816183115

839 96. Abere B, Li J, Zhou H, Toptan T, Moore PS, Chang Y. Kaposi's Sarcoma-Associated Herpesvirus-
840 Encoded circRNAs Are Expressed in Infected Tumor Tissues and Are Incorporated into Virions.
841 MBio. 2020;11. doi:10.1128/mBio.03027-19

842 97. Toptan T, Abere B, Nalesnik MA, Swerdlow SH, Ranganathan S, Lee N, et al. Circular DNA tumor
843 viruses make circular RNAs. Proc Natl Acad Sci U S A. 2018;115: E8737–E8745.
844 doi:10.1073/pnas.1811728115

845 98. Taylor JL, Bennett HN, Snyder BA, Moore PS, Chang Y. Transcriptional analysis of latent and
846 inducible Kaposi's sarcoma-associated herpesvirus transcripts in the K4 to K7 region. J Virol.
847 2005;79: 15099–106. doi:10.1128/JVI.79.24.15099-15106.2005 PMID: 16306581

848 99. Hu Z, Usherwood EJ. Immune escape of γ-herpesviruses from adaptive immunity. Rev Med Virol.
849 2014;24: 365–78. doi:10.1002/rmv.1791 PMID: 24733560

850 100. Lee H-R, Lee S, Chaudhary PM, Gill P, Jung JU. Immune evasion by Kaposi's sarcoma-associated
851 herpesvirus. Future Microbiol. 2010;5: 1349–65. doi:10.2217/fmb.10.105 PMID: 20860481

852 101. Wei F, Zhu Q, Ding L, Liang Q, Cai Q. Manipulation of the host cell membrane by human γ-
853 herpesviruses EBV and KSHV for pathogenesis. Virol Sin. 2016;31: 395–405. doi:10.1007/s12250-
854 016-3817-2 PMID: 27624182

855 102. Wei X, Lan K. Activation and counteraction of antiviral innate immunity by KSHV: an Update. Sci
856 Bull. 2018;63: 1223–1234. doi:10.1016/j.scib.2018.07.009 PMID: 30906617

857 103. Manners O, Murphy JC, Coleman A, Hughes DJ, Whitehouse A. Contribution of the KSHV and EBV
858 lytic cycles to tumourigenesis. Curr Opin Virol. 2018;32: 60–70. doi:10.1016/j.coviro.2018.08.014
859 PMID: 30268927

860 104. Forte E, Raja AN, Shamulailatpam P, Manzano M, Schipma MJ, Casey JL, et al. MicroRNA-
861 Mediated Transformation by the Kaposi's Sarcoma-Associated Herpesvirus Kaposin Locus. Hutt-
862 Fletcher L, editor. J Virol. 2015;89: 2333–2341. doi:10.1128/JVI.03317-14 PMID: 25505059

863 105. Gandy SZ, Linnstaedt SD, Muralidhar S, Cashman KA, Rosenthal LJ, Casey JL. RNA editing of the
864 human herpesvirus 8 kaposin transcript eliminates its transforming activity and is induced during
865 lytic replication. J Virol. 2007;81: 13544–51. doi:10.1128/JVI.01521-07 PMID: 17913828

866 106. Wang FZ, Akula SM, Pramod NP, Zeng L, Chandran B. Human herpesvirus 8 envelope glycoprotein

K8.1A interaction with the target cells involves heparan sulfate. J Virol. 2001;75: 7517–27. doi:10.1128/JVI.75.16.7517-7527.2001 PMID: 11462024

107. Birkmann A, Mahr K, Ensser A, Yağuboğlu S, Titgemeyer F, Fleckenstein B, et al. Cell surface heparan sulfate is a receptor for human herpesvirus 8 and interacts with envelope glycoprotein K8.1. J Virol. 2001;75: 11583–93. doi:10.1128/JVI.75.23.11583-11593.2001 PMID: 11689640

108. Akula SM, Pramod NP, Wang F-Z, Chandran B. Human Herpesvirus 8 Envelope-Associated Glycoprotein B Interacts with Heparan Sulfate-like Moieties. Virology. 2001;284: 235–249. doi:10.1006/viro.2001.0921 PMID: 11384223

109. Akula SM, Wang F-Z, Vieira J, Chandran B. Human Herpesvirus 8 Interaction with Target Cells Involves Heparan Sulfate. Virology. 2001;282: 245–255. doi:10.1006/viro.2000.0851 PMID: 11289807

110. Luna RE, Zhou F, Baghian A, Chouljenko V, Forghani B, Gao S-J, et al. Kaposi's Sarcoma-Associated Herpesvirus Glycoprotein K8.1 Is Dispensable for Virus Entry. J Virol. 2004;78: 6389–6398. doi:10.1128/JVI.78.12.6389-6398.2004 PMID: 15163732

111. Dollery SJ, Santiago-Crespo RJ, Chatterjee D, Berger EA. Glycoprotein K8.1A of Kaposi's Sarcoma-Associated Herpesvirus Is a Critical B Cell Tropism Determinant Independent of Its Heparan Sulfate Binding Activity. Longnecker RM, editor. J Virol. 2018;93. doi:10.1128/JVI.01876-18 PMID: 30567992

112. Li M, MacKey J, Czajak SC, Desrosiers RC, Lackner AA, Jung JU. Identification and characterization of Kaposi's sarcoma-associated herpesvirus K8.1 virion glycoprotein. J Virol. 1999;73: 1341–9. PMID: 9882339

113. Raab MS, Albrecht JC, Birkmann A, Yağuboğlu S, Lang D, Fleckenstein B, et al. The immunogenic glycoprotein gp35-37 of human herpesvirus 8 is encoded by open reading frame K8.1. J Virol. 1998;72: 6725–31. PMID: 9658120

114. Chandran B, Bloomer C, Chan SR, Zhu L, Goldstein E, Horvat R. Human Herpesvirus-8 ORF K8.1 Gene Encodes Immunogenic Glycoproteins Generated by Spliced Transcripts. Virology. 1998;249: 140–149. doi:10.1006/viro.1998.9316 PMID: 9740785

115. Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A. 1996;93: 14862–7. doi:10.1073/pnas.93.25.14862 PMID: 8962146

116. Deng J-H, Zhang Y-J, Wang X-P, Gao S-J. Lytic replication-defective Kaposi's sarcoma-associated herpesvirus: potential role in infection and malignant transformation. J Virol. 2004;78: 11108–20. doi:10.1128/JVI.78.20.11108-11120.2004 PMID: 15452231

117. Juillard F, Tan M, Li S, Kaye KM. Kaposi's Sarcoma Herpesvirus Genome Persistence. Front Microbiol. 2016;7: 1149. doi:10.3389/fmicb.2016.01149 PMID: 27570517

118. Purushothaman P, Dabral P, Gupta N, Sarkar R, Verma SC. KSHV Genome Replication and Maintenance. Front Microbiol. 2016;7: 54. doi:10.3389/fmicb.2016.00054 PMID: 26870016

119. Lallemand F, Desire N, Rozenbaum W, Nicolas JC, Marechal V. Quantitative analysis of human herpesvirus 8 viral load using a real-time PCR assay. J Clin Microbiol. 2000;38: 1404–8. PMID: 10747115

120. Chiu Y-F, Sugden AU, Fox K, Hayes M, Sugden B. Kaposi's sarcoma-associated herpesvirus stably clusters its genomes across generations to maintain itself extrachromosomally. J Cell Biol.

2017;216: 2745–2758. doi:10.1083/jcb.201702013 PMID: 28696226

121. Xiao K, Yu Z, Li X, Li X, Tang K, Tu C, et al. Genome-wide Analysis of Epstein-Barr Virus (EBV) Integration and Strain in C666-1 and Raji Cells. J Cancer. 2016;7: 214–24. doi:10.7150/jca.13150 PMID: 26819646

122. Zaldumbide A, Ossevoort M, Wiertz EJHJ, Hoeben RC. In cis inhibition of antigen processing by the latency-associated nuclear antigen I of Kaposi sarcoma Herpes virus. Mol Immunol. 2007;44: 1352–1360. doi:10.1016/j.molimm.2006.05.012 PMID: 16828498

Figure Legends

Figure 1. KSHV genomes in BCBL-1 cells have low point mutational diversity. (A) Schematic representation of a linear KSHV genome, with genes colored in green and the major repeat regions in orange. The locations of the K1, vIL-6, vBCL-2, RTA, LANA and K15 genes used for genome quantitation are indicated. (B) Raw (light blue), sUMI-consensus (blue) and dUMI-consensus (dark blue) read coverage along the *de novo* assembled, BCBL-1 KSHV genome. Major repeat regions were masked (gray columns). (C) Bubble plot of minor sequence variants. Each bubble represents a position within the genome at which a variant base or indel was detected, colored by whether they were predicted to be silent or protein-altering mutations. Mutations likely to be silent included synonymous and intergenic point mutations, while protein-altering mutations included non-synonymous, nonsense and frameshift mutations. Bubble height represents variant base frequency among dUMI-consensus reads at that position. Vertical grey columns represent the masked repeat regions.

Figure 2. Workflow for analyzing intra-host KSHV genome diversity from clinical samples. Each study participant contributed KS tumors and oral swabs. Sequencing libraries were prepared from DNA extracts from each sample with adaptors containing duplex Unique Molecular Identifiers (dUMIs). Adaptor-labelled DNA libraries were enriched using biotinylated RNA baits homologous to KSHV sequences. Captured DNA was PCR-amplified to levels sufficient for Illumina HiSeq sequencing. For most samples, libraries were subjected to a second round of enrichment and PCR amplification. Upon sequencing, whole KSHV genomes were first assembled *de novo* from each sample without the use of dUMIs. The sample-specific genomes generated (sample-consensus) were then used as reference to map the consensus of reads with identical dUMI-tags (i.e., dUMI-consensus reads).

941

942 **Figure 3. Point mutational diversity in KSHV genomes from tumors and oral swabs.** Bubble plots of
943 minor sequence variants remaining after removal of PCR errors, in KSHV genomes from tumors (A) and
944 oral swabs (B). Each bubble represents a variant base or indel, colored by whether they were predicted
945 to be silent or protein-altering mutations. Silent mutations include synonymous and intergenic point
946 mutations, while protein-altering mutations included non-synonymous, nonsense and frameshift
947 mutations. Hollow circles represent mutations occurring in homopolymer runs. Bubble heights
948 represent the frequency of the variant base among dUMI-consensus reads at that position. Vertical gray
949 columns represent the masked repeat regions. The region containing the K1 gene is indicated with
950 arrows at the bottom of the figure.

951

952 **Figure 4. KSHV genomes in the U003-C tumor harbor a deletion within the K8.1 gene.** (A) Read
953 coverage of the U003-C KSHV genome, showing a 6-bp gap (red arrow) where no read pairs were
954 mapped. (B) Cartoon of the *de* novo-assembly sequences generated at either side of the gap, both
955 ended within the K8.1 gene intron and continued into terminal repeat (TR) sequences. Green and yellow
956 arrows show the directions of the K8.1 gene and terminal repeat sequences, respectively. Blue arrows
957 show the position of PCR primers used to confirm breakpoint junctions, with the expected PCR product
958 sizes. (C) PCR products generated from U003-C tumor DNA using primers flanking the gap. The 443-bp
959 PCR product expected if the K8.1 gene intron was intact was not detected from U003-C (left column),
960 whereas the expected band was detected in tumor U007-B (right column) from another person. (D)
961 Hemi-nested PCR of U003-C tumor DNA for the K8.1-TR (left) and TR-K8.1 (right) junctions produced
962 products of the predicted sizes. These structures were confirmed by Sanger sequencing (data not
963 shown). No K8.1-TR or TR-K8.1 junction fragment was produced from BCBL-1 DNA. The light bands at

964 the TR-K8.1 lane under BCBL-1 were found to be amplicons generated from the forward primer
965 sequence (indicated with * in panel B) overlapping with K8.1; this primer was used since the rest of the
966 connected TR sequence was GC-rich and unsuitable for primer design.

967

968 **Figure 5. KSHV genomes in two tumors from participant U008 had a 14.8-kb region flanking Internal**
969 **Repeat 1 (IR1) duplicated and translocated to Internal Repeat 2 (IR2).** Total, sUMI and dUMI-consensus
970 read coverage of tumor B (A) and D (B) genomes from individual U008. (C) Annotations of the region
971 with 1.5-2X read coverage, with genes in green, repeat regions in orange, and long non-coding RNAs in
972 red. Many reads on the edge of this region continue into IR2 (red arrows). Annotations are from the
973 KSHV reference strain GK18. (D) Cartoon showing the duplication of the 14.8 kb region into IR2 and the
974 PCR primers used to examine the genomic rearrangement in unsheared tumor DNA extracts from
975 tumors U008-B and U008-D. PCR products produced from primer pairs numbered in D from U008-B and
976 BCBL-1 (E) and in U008-D (F). All visible bands were excised from the agarose gel and sequenced,
977 confirming the junction sequences. Primer pairs # 9-12 produced no PCR products discernible on an
978 agarose gel and are not shown here.

979

980 **Figure 6. KSHV genomes in U020-B and U020-C have large, distinct deletions.** Total and dUMI-
981 consensus read coverages of U020-B (A) and U020-C (B) KSHV genomes. GK18 reference annotations in
982 the high-coverage regions of U020-B (C) and U020-C (D). (E) Cartoon showing the region encompassing
983 the high coverage region of U020-C viral genomes, leaving a 16.5-kb region connected to TR. (F) PCR
984 primers used to examine the deletions. Primers to unique genomic sequences are in blue, while primers
985 to repeat sequences are in orange. (G) PCR products produced from primer pairs numbered in E, with
986 DNA from U020-C and BCBL-1 as templates. Bands from lanes 1, 2 and 5 were excised from the agarose

gel and sequenced. Attempts to sequence the light bands in U020-C #7 and #9 were unsuccessful. Row 11 primers, in which the forward primer binds to unique genomic sequences preceding the TR, yielded no discernible product (not shown).

Figure 7. KSHV genome structures in participant U003. Junction sequences marking the genomic aberration in U003-C were detected in all 4 other tumors tested, while intact K8.1 sequences were detected in only 2. The cartoon shows the breakpoints in the K8.1 intron of U003-C extending into TR sequences, along with PCR primers used to confirm the genome structure. PCR products from other tumors of participant U003 and from the BCBL-1 cell line are shown. All visible bands were excised from the agarose gel and their structures confirmed by Sanger sequencing.

Figure 8. Schematic representation of the 5 aberrant KSHV genomes discovered in KS tumors. Specific details and evidence for each are referred to in the text and in Table 3.

S1 Figure. dUMI-adaptors and primers for duplex sequencing. During library preparation, sheared DNA fragments were A-tailed and ligated with forked, double-stranded oligonucleotides containing Illumina TruSeq universal adaptor sequences, 12-random base pairs as dUMI and spacer sequences. The adapted DNA libraries were PCR amplified before enrichment with primers mws13 and mws20, which bind to Illumina Truseq adaptors. Primer mws21 containing sample index ID for multiplex sequencing was used for PCR following enrichment. DNA libraries post processing are shown at the bottom.

S2 Figure. Workflow of genome assembly and variant analysis. KSHV genomes were first assembled *de*

novo from sequence reads of each sample, before being used as reference for mapping their respective dUMI-consensus reads. See details in the Methods section. Discrepancies in bases between the sample-consensus genome and mapped dUMI-consensus reads were taken to be real intra-sample variants.

S3 Figure. Read coverage. Raw (light blue), sUMI (blue) and dUMI-consensus (dark blue) read coverage in log scale along the de novo assembled, sample-consensus KSHV genomes in tumors (A) and oral swabs (B) examined in this study. Major repeat regions were masked with inserted Ns in all sample-consensus genomes.

S4 Figure. Potential sequencing artifacts. (A) KSHV intrasample variant frequency as a function of read coverage. Sample variant frequencies were estimated (Table 1) when at least 100 viral genomes were sampled. (B) Minor variants detected in dUMI-consensus reads of all samples, by type of base substitution.

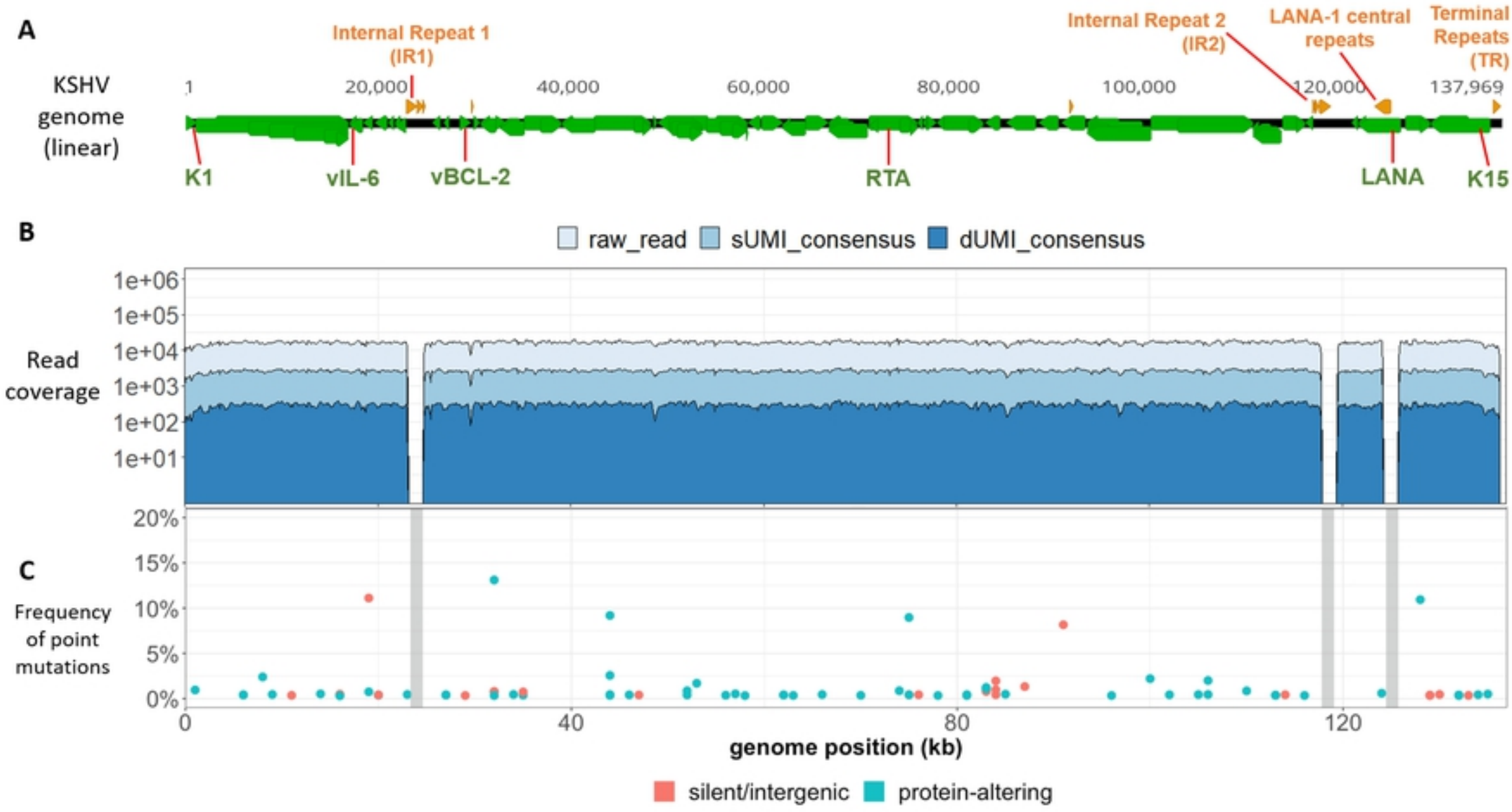
S5 Figure. KSHV phylogenetic relationships by variable regions K1 and K15 and by whole genomes. Phylogenetic trees of (A) K1 genes, (B) K15 genes and (C) whole genomes from this study and of select genomes from other publications. K1 and K15 subtypes are indicated to the right of the K1 (A) and K15 (B) trees. (D) A neighbor-net phylogenetic network of all published KSHV genomes to date, color-coded by genome types proposed in [36]: P1 in green, P2 in blue, N in purple, M1 in red and M2 in maroon. All *de novo*-assembled genomes from this study are in bolded italics.

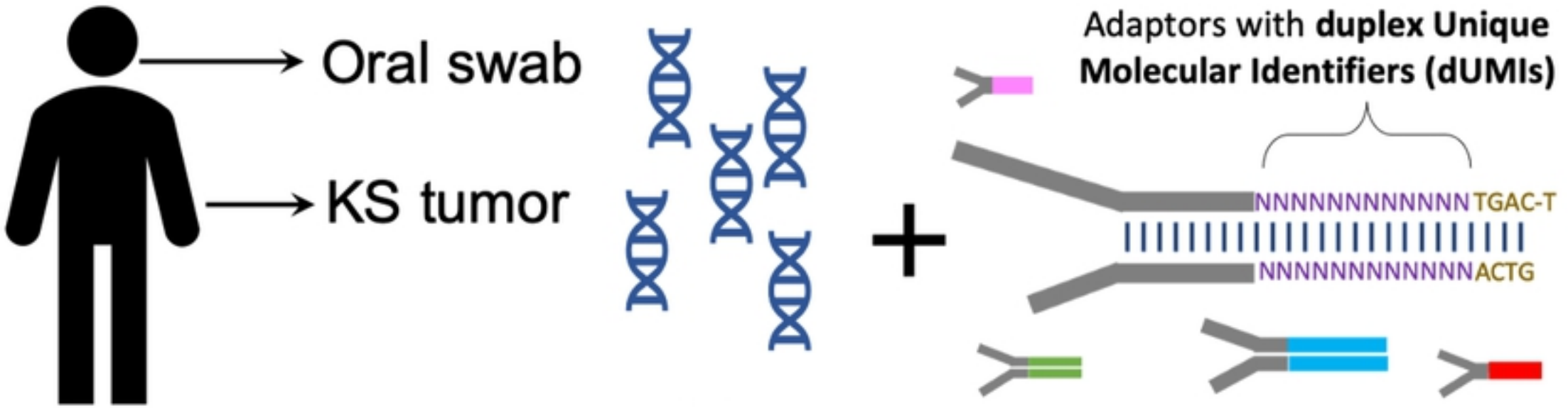
S6 Figure. U008-B and U008-D were from distinct lesions on the left leg. U008-B biopsy was obtained

from lesions in the upper thigh, while U008-D was biopsied from a large lesion on the knee.

S7 Figure. Mutations of KSHV genomes in tumors from participants 004 and 003. (A) Alignment of KSHV genomes from Participant 004, showing a 28-bp deletion in the K8.1 promoter in U004-D. U004-D and U004-C are from tumors while U004-o1 is from an oral swab. (B) The only intra-host synonymous mutation found in this study, within miR-K10 in participant 003. (C) Sequence chromatograms of miR-K10 in other tumors of participant U003, with a T in all tumors and a mixture of T and the database consensus C in a minority of viruses in U003-G.

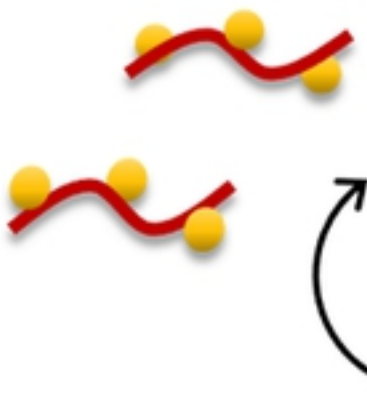
S8 Figure. Scaffolds of EBV sequences produced in 4 KS tumors map to the EBNA-1 gene. A portion of the EBV genome is illustrated with an example of the 2 scaffolds (grey striped bars) generated from the reads of 4 KS tumors. Vertical stripes inside the gray bars indicate mismatches to the EBV reference (GenBank ID: DQ279927).

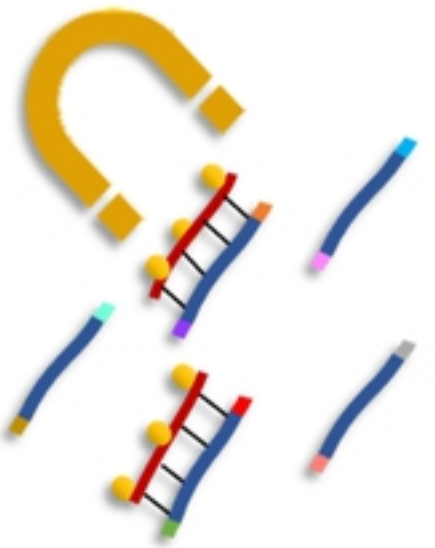

A
KSHV genome (linear)
1
20,000
40,000
60,000
80,000
100,000
120,000
137,969
Internal Repeat 1 (IR1)
Internal Repeat 2 (IR2)
LANA-1 central repeats
Terminal Repeats (TR)
K1
vIL-6
vBCL-2
RTA
LANA
K15
B
raw_read
sUMI_consensus
dUMI_consensus
Read coverage
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
C
Frequency of point mutations
20%
15%
10%
5%
0%
0
40
80
120
genome position (kb)
silent/intergenic
protein-altering



PCR

Enrich for KSHV sequences up to twice using biotinylated RNA baits





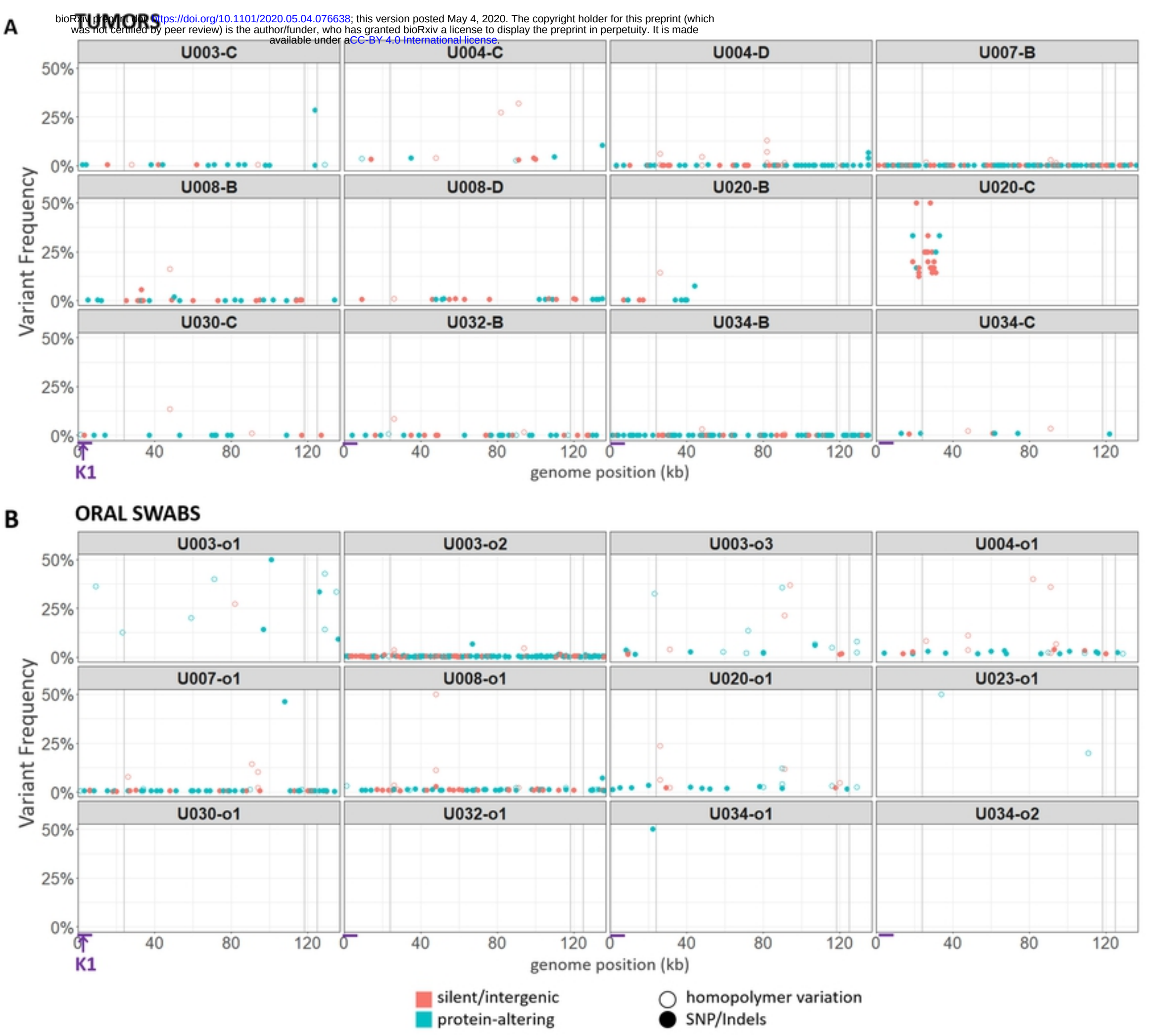
PCR

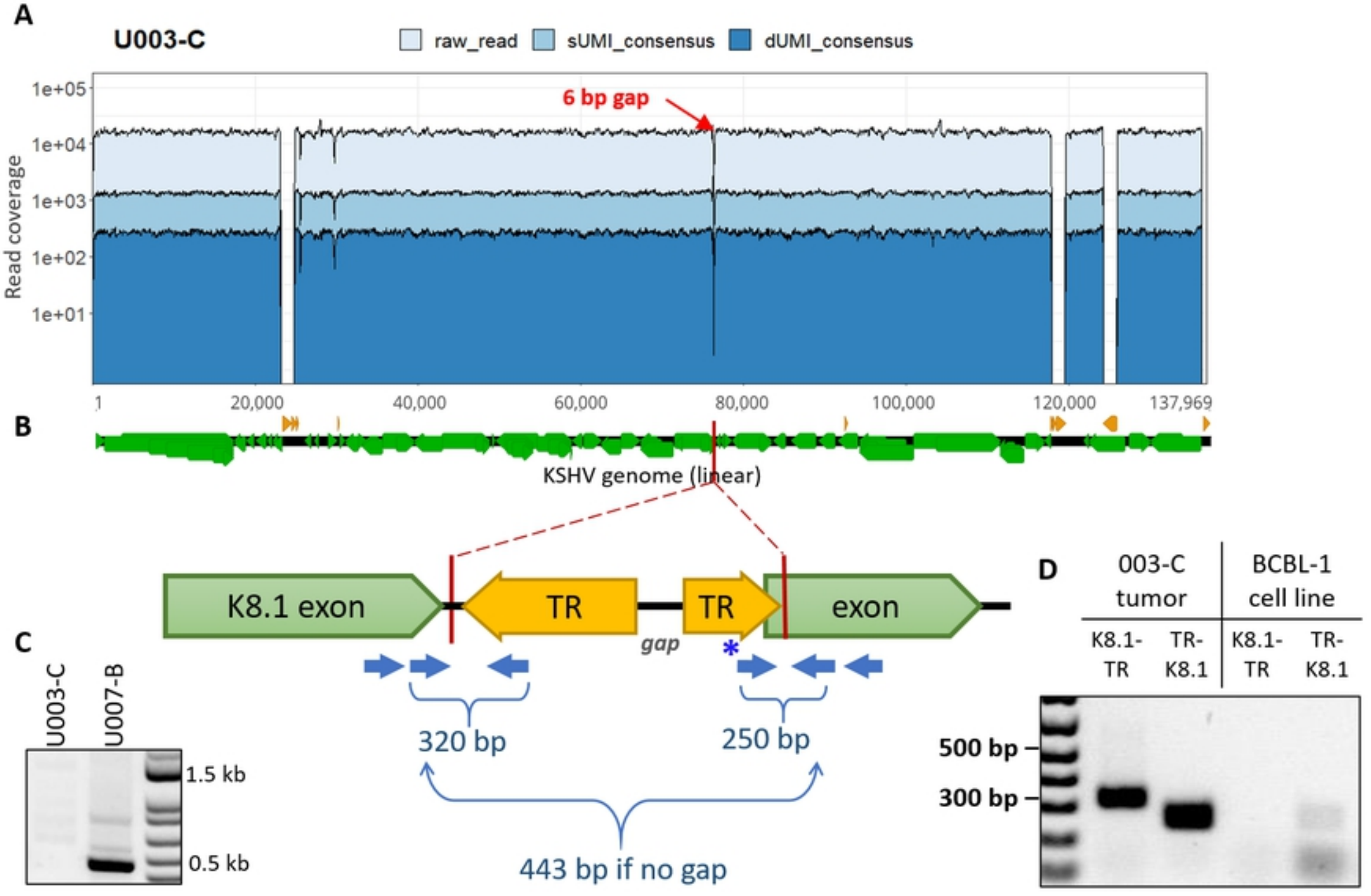
Illumina HiSeq

Assemble *de novo* a <u>sample-consensus</u> KSHV genome

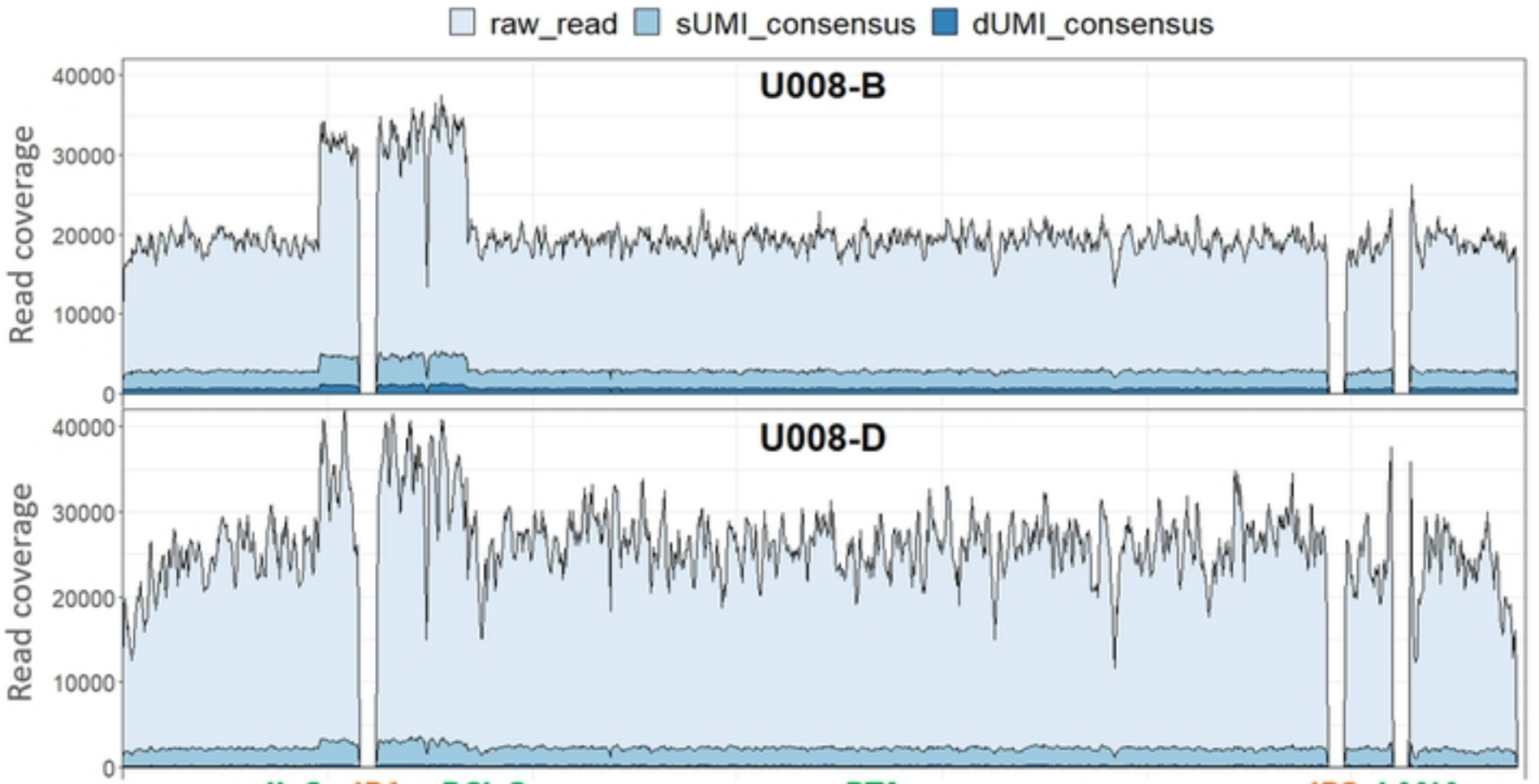
Use as reference

Map dUMI-consensus reads



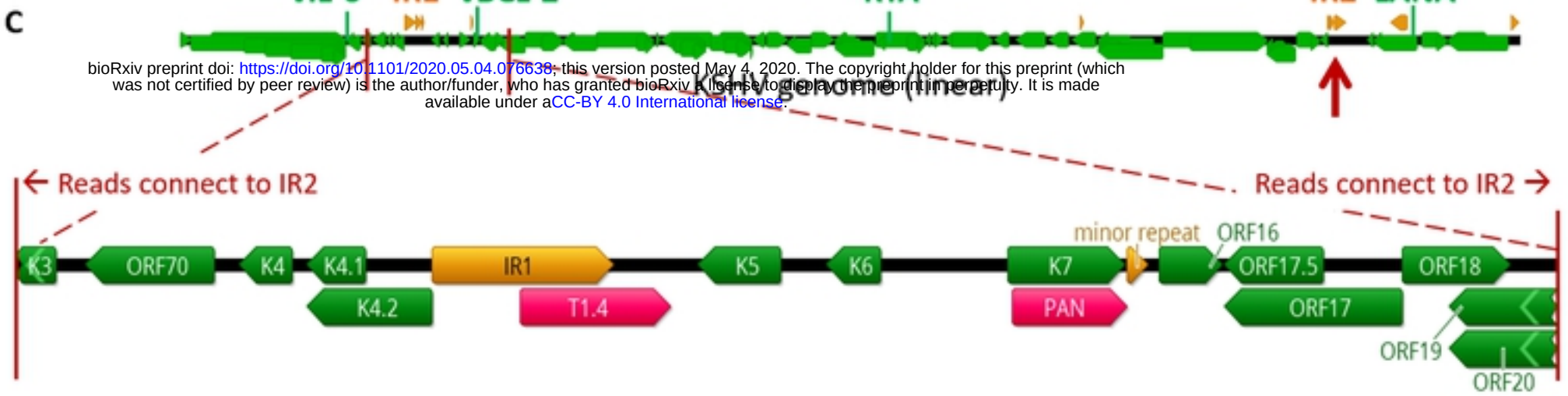

A
U003-C
raw_read
sUMI_consensus
dUMI_consensus
6 bp gap
Read coverage
1e+05
1e+04
1e+03
1e+02
1e+01
1
20,000
40,000
60,000
80,000
100,000
120,000
137,969
B
KSHV genome (linear)
K8.1 exon
TR
TR
exon
gap
*
C
U003-C
U007-B
1.5 kb
0.5 kb
320 bp
250 bp
443 bp if no gap
D
003-C tumor
BCBL-1 cell line
K8.1-TR
TR-K8.1
K8.1-TR
TR-K8.1
500 bp
300 bp

A

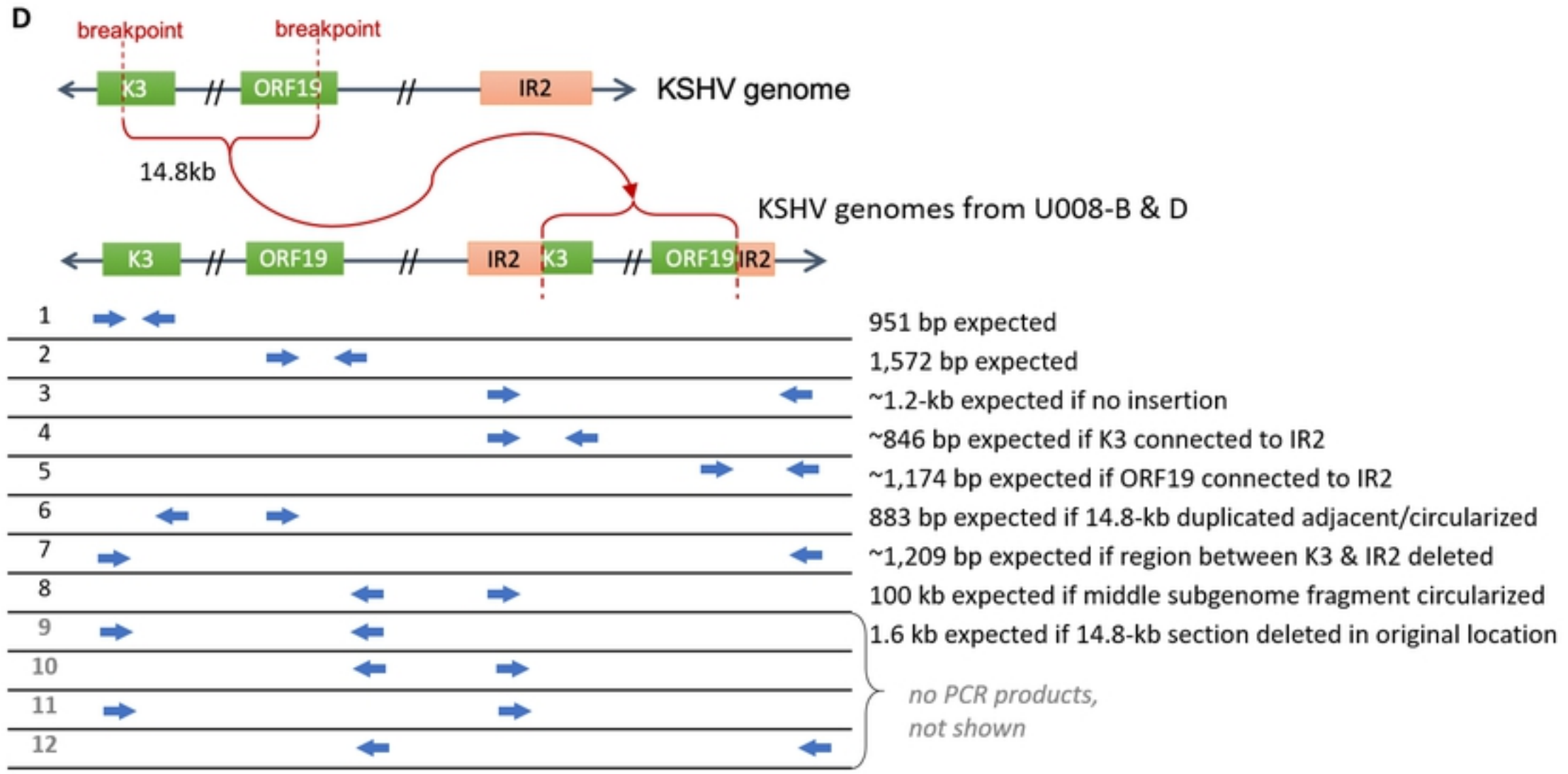


B

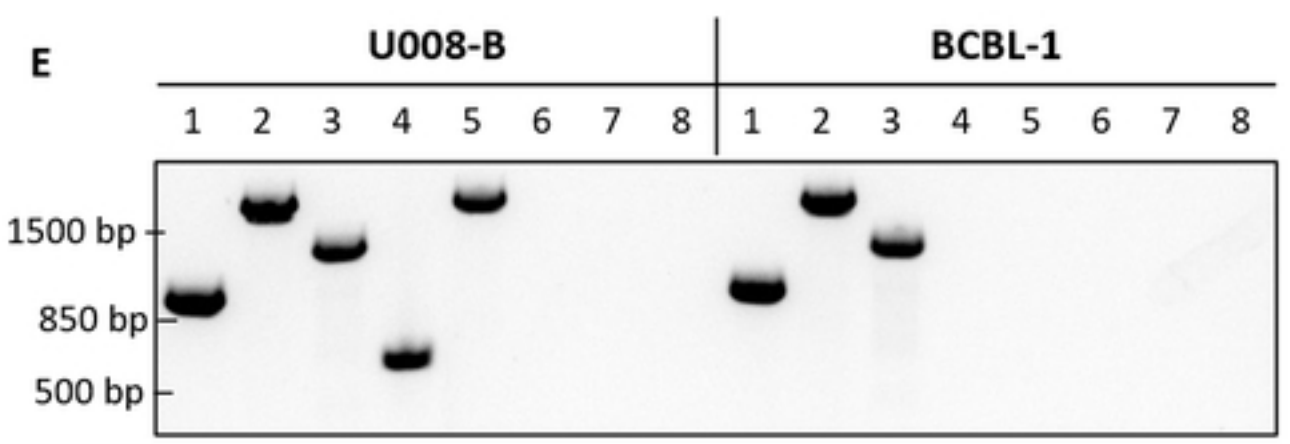
C



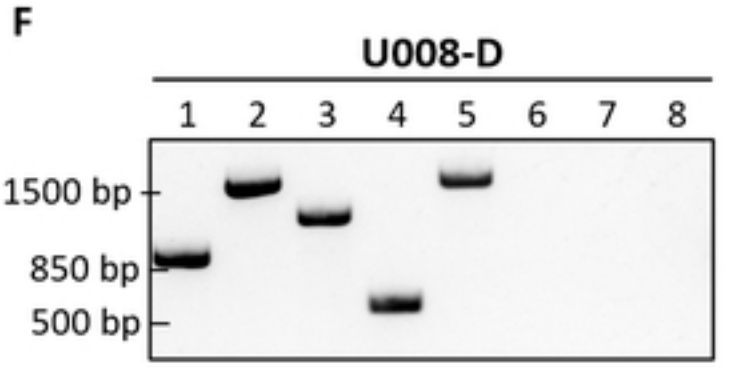
D

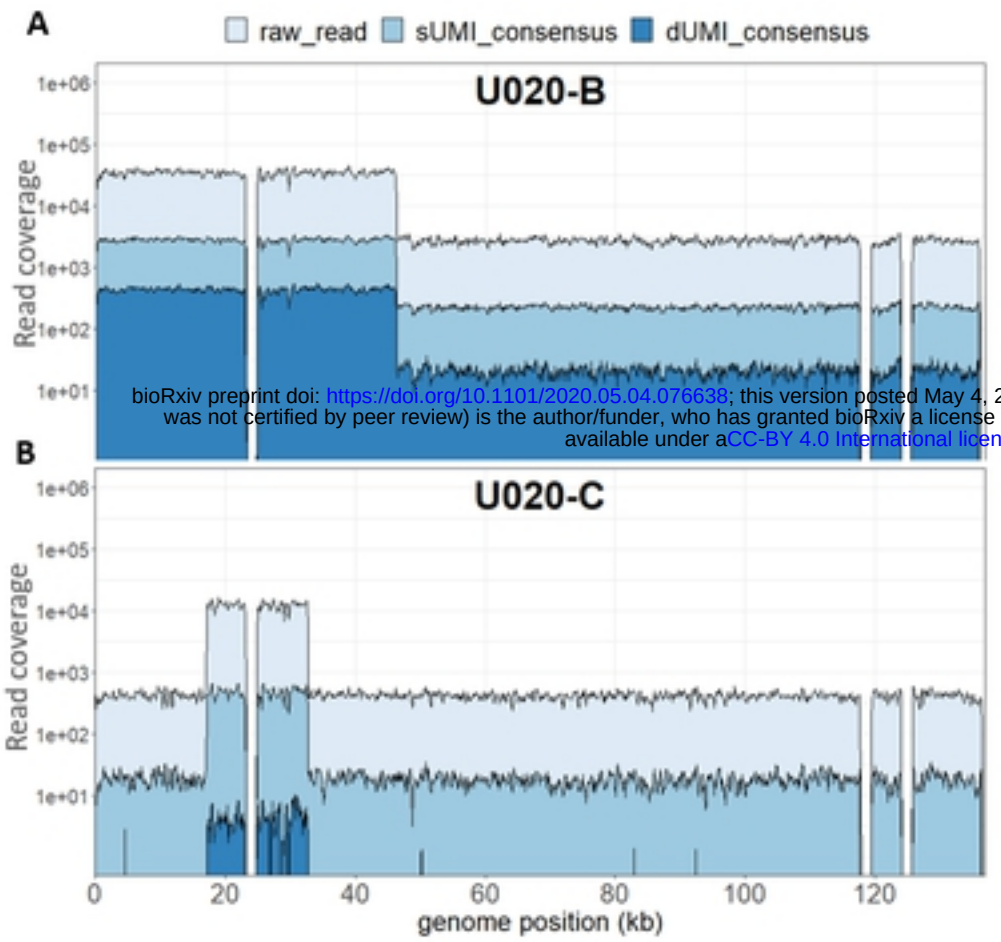


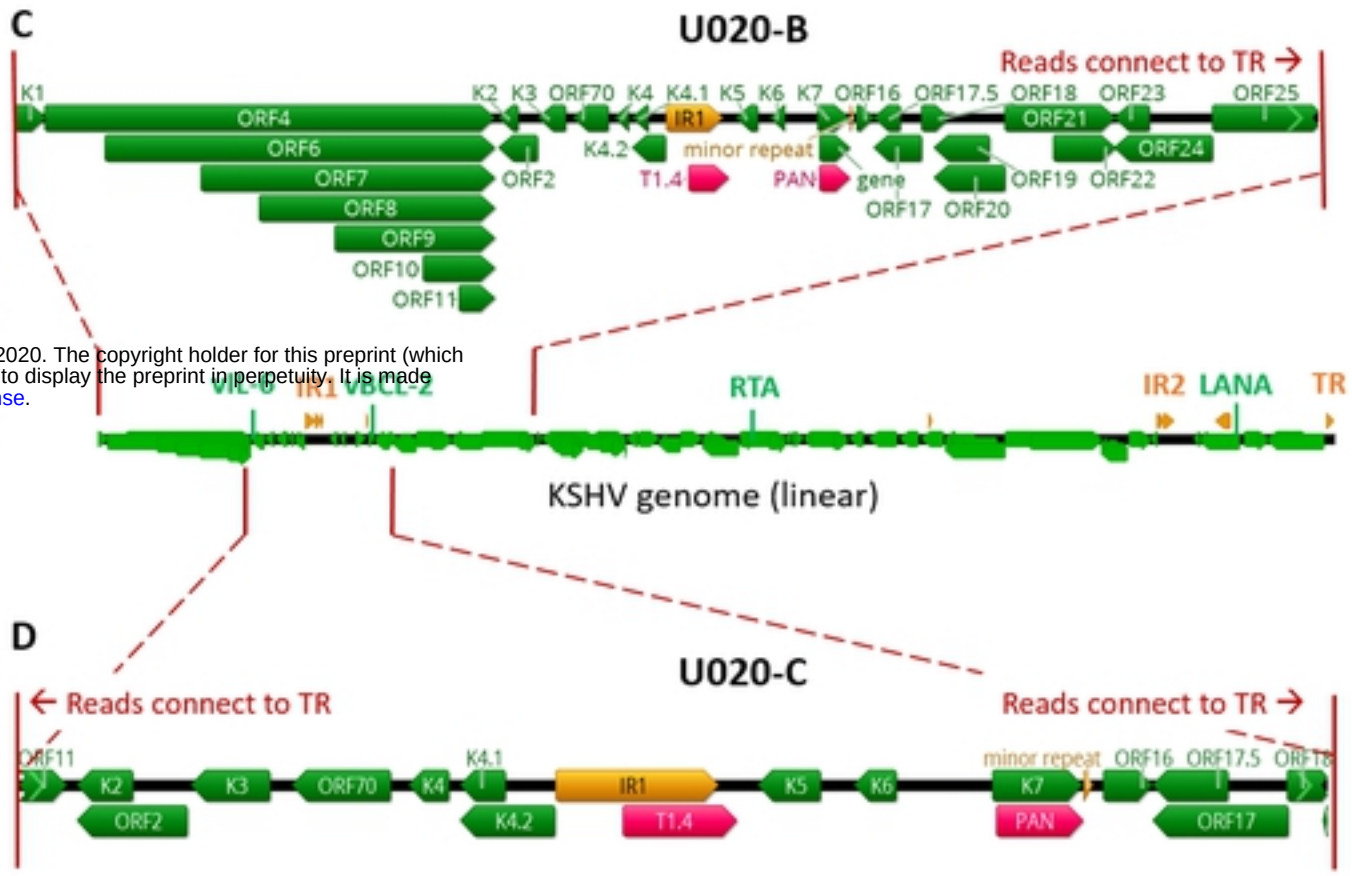
E

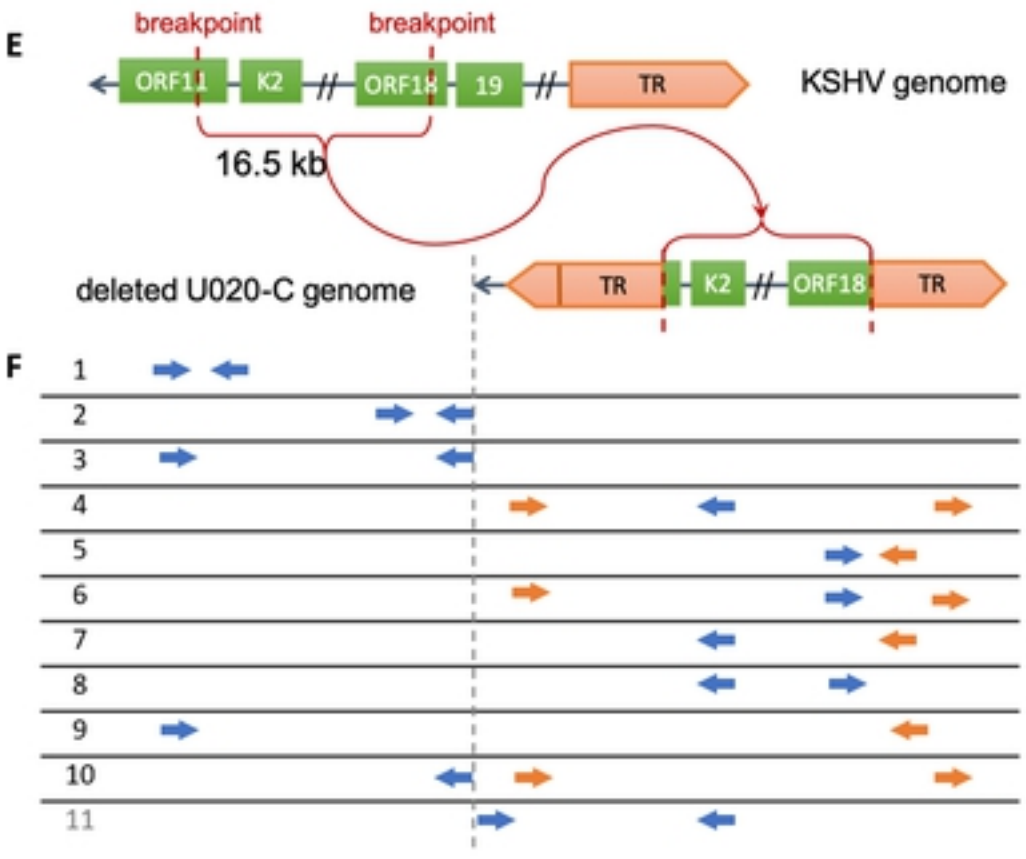


F

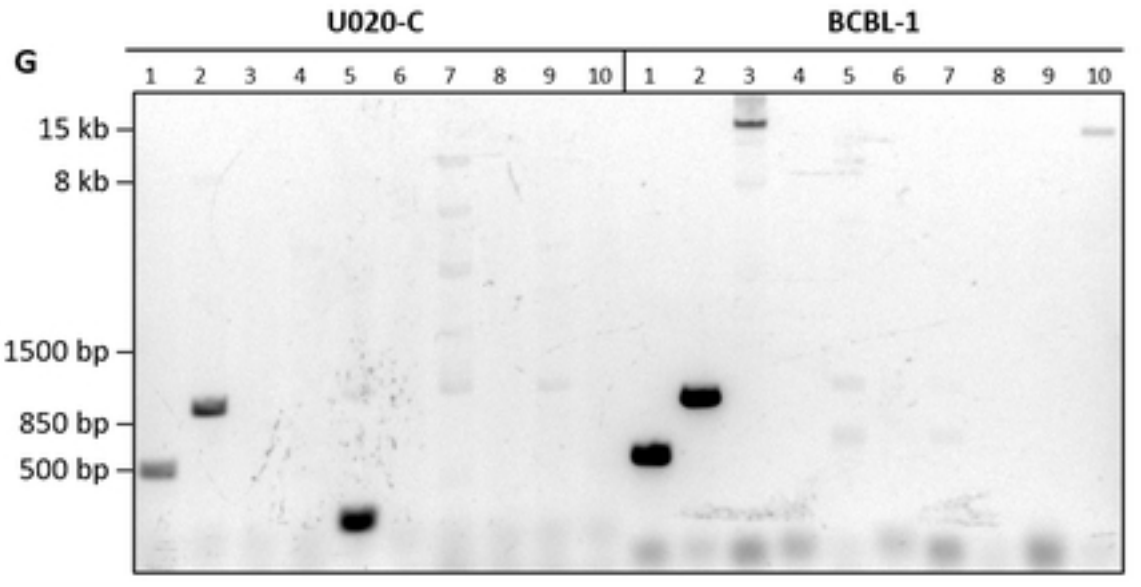


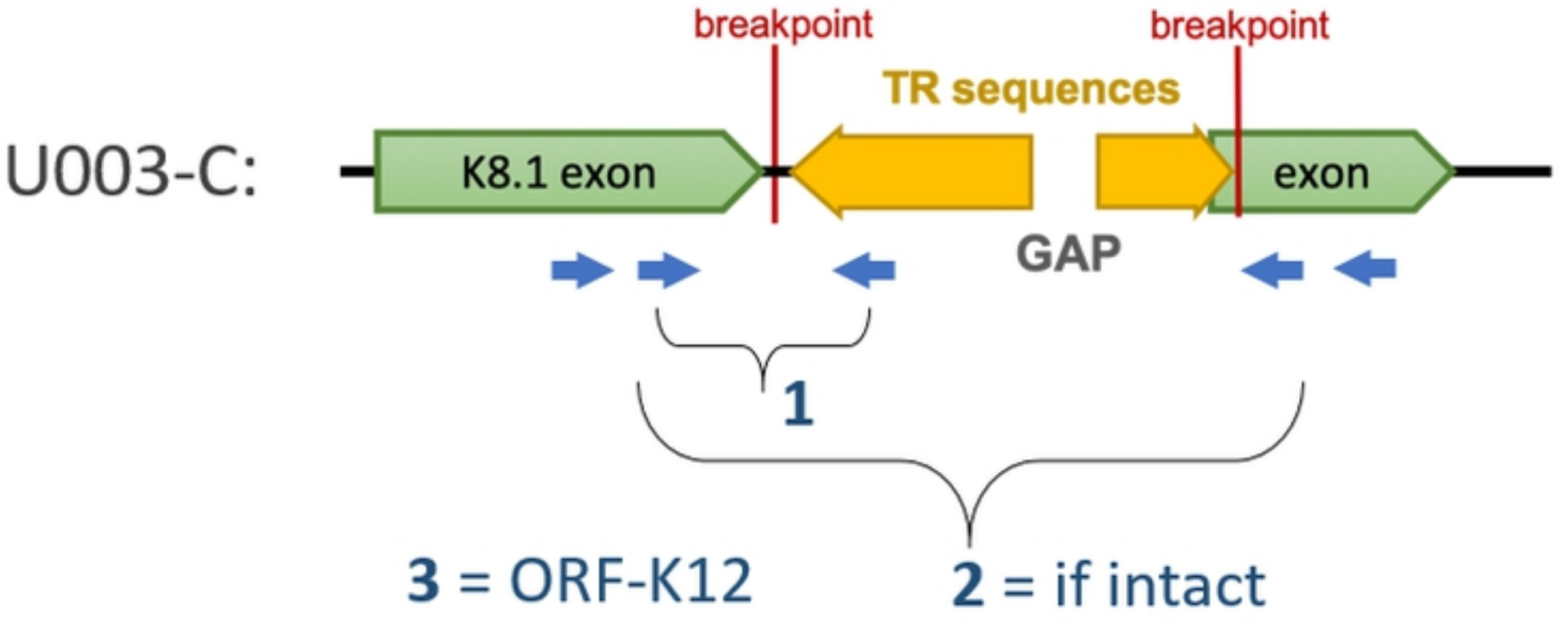


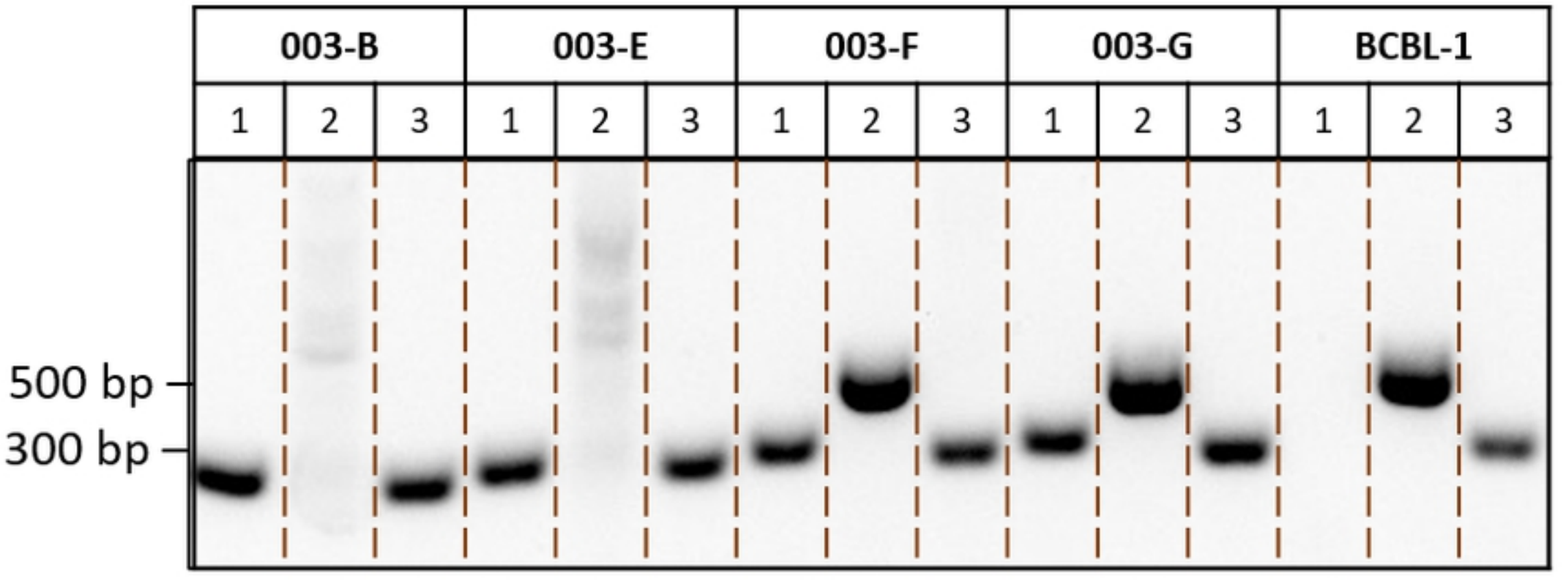


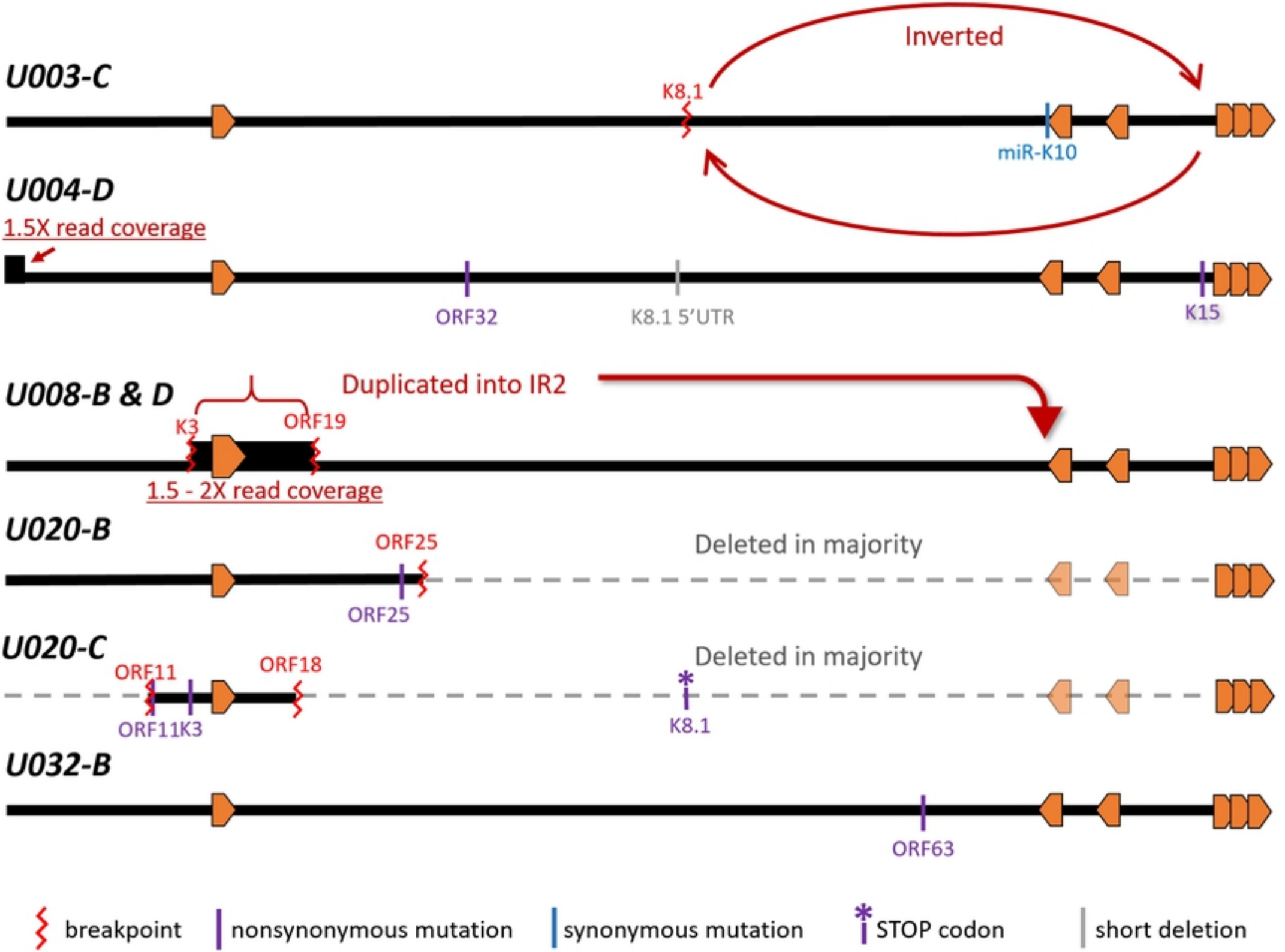


478 bp expected in wild-type
803 bp expected in wild-type
892 bp if 16.5-kb section deleted from original location
at least ~1 kb if K2 connected to TR (will include 1-unit length TR)
240 bp if ORF18 connected to TR
at least ~1 kb if ORF18 connected to TR in reverse
277 bp if K2 connected to TR in reverse
392 bp if 16.5-kb region duplicated adjacent/circularized
328 bp if ORF11 connected to TR (end of ORF11 to TR deleted)
at least ~1 kb if ORF 18 end connected to TR
~1 kb if unique sequence upstream of TR connected to ORF11 fragment (no band, not shown)




U003-C:
breakpoint
breakpoint
TR sequences
K8.1 exon
exon
GAP
1
2
3 = ORF-K12
2 = if intact


003-B
003-E
003-F
003-G
BCBL-1
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
500 bp
300 bp


U003-C
Inverted
K8.1
miR-K10
U004-D
1.5X read coverage
ORF32
K8.1 5'UTR
K15
U008-B & D
Duplicated into IR2
K3
ORF19
1.5 - 2X read coverage
U020-B
ORF25
Deleted in majority
ORF25
U020-C
ORF11
ORF18
Deleted in majority
ORF11K3
K8.1
U032-B
ORF63
breakpoint
nonsynonymous mutation
synonymous mutation
STOP codon
short deletion