Copy number heterogeneity, large origin tandem repeats, and interspecies recombination in HHV-6A and HHV-6B reference strains

Alexander L. Greninger[1,2,#], Pavitra Roychoudhury[1,2], Negar Makhsous[1], Derek Hanson[3], Jill Chase[5], Gerhard Krueger[6], Hong Xie[1], Meei-Li Huang[1], Lindsay Saunders[1], Dharam Ablashi[5], David M. Koelle[1,2,3,4], Linda Cook[1], Keith R. Jerome[1,2]

[1]Department of Laboratory Medicine, University of Washington, Seattle, WA, USA

[2]Fred Hutchinson Cancer Research Institute, Seattle, WA, USA

[3]Department of Medicine, University of Washington, Seattle, WA, USA

[4]Department of Global Health, University of Washington, Seattle, WA, USA

[5]HHV-6 Foundation, Santa Barbara, CA, USA

[6]Department of Pathology and Laboratory Medicine, University of Houston, Houston, TX, USA

[#]Corresponding author: agrening@uw.edu

Abstract

Quantitative PCR is the diagnostic pillar for clinical virology testing, and reference materials are necessary for accurate, comparable quantitation between clinical laboratories. Accurate quantitation of HHV-6 is important for detection of viral reactivation and inherited chromosomally integrated HHV-6 in immunocompromised patients. Reference materials in clinical virology commonly consist of laboratory-adapted viral strains that may be affected by the culture process. We performed next-generation sequencing to make relative copy number measurements at single nucleotide resolution of eight candidate HHV-6A and seven HHV-6B reference strains and DNA materials from the HHV-6 Foundation and Advanced Biotechnologies. 11 of 17 (65%) HHV6 candidate reference materials showed multiple copies of the origin of replication upstream of the U41 gene by next-generation sequencing. These large tandem repeats arose independently in culture-adapted HHV-6A and HHV-6B strains, measuring 1254 bp and 983 bp, respectively. Copy number measured between 4-10X copies relative to the rest of the genome. We also report the first interspecies recombinant HHV-6 strain with a HHV-6A GS backbone and >5.5kb region from HHV-6B Z29 from U41-U43 that covered the origin tandem repeat. Specific HHV-6A reference strains demonstrated duplication of regions at UL1/UL2, U87, and U89, as well as deletion in the U12-U24 region and U94/95 genes. HHV-6 strains derived from cord blood mononuclear cells from different labs on different continents revealed no copy number differences throughout the viral genome. These data indicate large origin tandem duplications are an adaptation of both HHV-6A and HHV-6B in culture and show interspecies recombination is possible within the *Betaherpesvirinae*.

Importance:

Anything in science that needs to be quantitated requires a standard unit of measurement. This includes viruses, for which quantitation increasingly determines definitions of pathology and guidelines for treatment. However, the act of making standard or reference material in virology can alter its very usefulness through genomic duplications, insertions, and rearrangements. We used deep sequencing to examine candidate reference strains for HHV-6, a ubiquitous human virus that can reactivate in the immunocompromised population and is integrated into the human genome in every cell of the body for 1% of people worldwide. We found large tandem repeats in the origin of replication for both HHV-6A and HHV-6B that are selected for in culture. We also found the first interspecies recombinant between HHV-6A and HHV-6B, a phenomenon that is well-known in alphaherpesviruses but to date has not been seen in betaherpesviruses. These data critically inform HHV-6 biology and the standard selection process.

Keywords: human herpesvirus 6, HHV-6, HHV-6A, HHV-6B, copy number, secondary standards, Z29, GS, origin of replication, origin, tandem repeat, nanopore, nanopore sequencing, direct repeat, quantitative PCR

Introduction

Human herpesvirus 6A and 6B (HHV-6) are ubiquitous human viruses with human exposure levels >90% by the age of 2 years old as measured by serological assays performed worldwide (1, 2). Both HHV-6A and HHV-6B establish chronic infections in the majority of infected individuals, leading to asymptomatic persistent viral shedding (3, 4). Exanthema subitum is the most common HHV-6 related infection seen after a primary exposure in 6 month to 3-year old children. Less frequently the virus can result in seizures, gastrointestinal and respiratory symptoms, thrombocytopenia, hepatitis, colitis, and CNS infections. Additionally, both HHV-6A and HHV-6B have been shown to integrate into host chromosomes in the telomere regions and be passed from parents to their children as inherited chromosomally integrated HHV-6 (iciHHV-6). The potential to measure HHV-6 from genomic DNA from these patients, as well as possible reactivation from the integrated HHV-6 make the diagnosis of HHV-6 infection from serum or plasma viral load testing challenging.

Accurate and sensitive real-time PCR assays that detect and quantify HHV-6 are critical to diagnosis and monitoring of the variety of manifestations of HHV-6 associated disease. Recently several quantitative cut-offs have been proposed that are associated with end-organ disease or iciHHV-6 status (5–7). There have been only a few limited studies comparing PCR methods between clinical labs (8–10). In addition, newer studies often utilize commercial reference laboratories or commercial reagents that in general do not make their primer/probe locations known. Of published studies where the PCR locations are described, many areas of the genome have been used, with no consistent location chosen. A review of the PCR methods used in 46 recent published papers (years

2014-2017) revealed the use of 17 different primer sets at multiple locations throughout the genome (at U6, 12, 13, 22, 27, 31, 32, 38, 41, 57, 65, 66, 67, 69, 90, 95, 100), with only the U31 and U65-66 primers used more than twice. Not surprisingly, this lack of consistency has contributed to a significant lack of consistency in test results as measured in cross-lab proficiency testing where quantitative differences as high as 4 logs have been seen (10, 11). These studies hint that between-lab results may be improved if diagnostic testing was done with a more limited number of high-performing primer sets. Finally, effective primer designs have been significantly limited by the lack of available DNA sequences.

Previous studies have identified the critical role that standardized materials play in the ability to establish clinical viral load cut-offs, establish assay sensitivity, and compare results between laboratories (10, 12–16). Efforts are currently underway at the National Institute for Biological Standards and Controls (NIBSC) in the United Kingdom to prepare WHO international standard material for both HHV-6A and 6B. However, since this effort utilizes cultured HHV-6 reference strains, it is unknown how well the materials utilized will reflect sequences in clinical isolates. Given the wide range of primer set locations used in labs, it is critical that the entire genome of any reference materials be studied

HHV-6A and -6B genomes have approximately 90% nucleotide identity to each other and about 50% similarity with the closest-related human betaherpesvirus, HHV-7. The genome is approximately 160-170 kb and contains many of the gene and regulatory elements present in the genome of other betaherpesviruses. Several prototypic strains of HHV-6 have been identified and utilized including GS, U1102, SIE, LHV, Z29, and

HST. Until very recently, there were fewer than 200 HHV-6 sequences in Genbank, including only 3 complete genomes. Recently, two large scale efforts have sequenced more than 150 near-full length HHV-6B genomes from four continents (17, 18).

In an effort to determine whether the available cultured "reference" strains have undergone significant changes during culture similar to that seen with the recently produced WHO BK and JC strains, we obtained 15 strains from the HHV-6 Foundation repository and used shotgun sequencing to obtain full-length genomes and estimate of copy number (19, 20). We found 9 of the 15 strains had high-copy tandem repeat amplifications in the origin of replication, including the first described origin amplification in an HHV-6A strain. We also describe the first HHV-6A/HHV-6B interspecies recombinant. Other HHV-6 reference materials had multiple loci with copy number variation of up to 20X. These duplications, deletions, and rearrangements may impact the utility of these strains for the production of standard materials for PCR testing. Changes in the genome of these strains in culture may have impact on the results of current and future studies utilizing these materials.

Materials and Methods

HHV-6 reference strains and DNA materials

HHV-6 culture isolates were obtained from the HHV-6 Foundation. The original HHV-6 isolate GS belongs to HHV-6A and was first isolated at the National Cancer Institute, NIH in 1986 from an AIDS patient. The HHV-6A GS strain was grown in HSB2, which is a human T-cell leukemic line derived from the peripheral blood of a child. The GS early passage isolate obtained from the HHV-6 Foundation is a low-

127 passaged HHV-6A GS isolate that was only passaged 4 times in cord blood mononuclear
128 cells (CBMC). The GS early passage isolate was sequenced in 2013 after a brief
129 expansion in CBMC and sequencing reads were obtained from L. Flamand (21). The
130 HHV-6A DA strain was isolated at the NCI from a patient with chronic fatigue syndrome
131 and was grown in the HSB2 cell line. The HHV-6A CO strains CO1, CO2, CO3, CO4,
132 CO7 were isolated from patients with collagen vascular diseases including systemic
133 lupus erythematosus, atypical polyclonal lymphoproliferation, rheumatoid arthritis, and
134 unclassified collagen vascular disease (22). The HHV-6A SIE strain was isolated from
135 an HIV-positive leukemia patient from the Ivory Coast and grown in PHA-stimulated
136 CBMCs. The HHV-6B strains HST, KYO, ENO, and NAK were isolated from Japanese
137 patients with exanthema subitum in 1988 (23). The HHV-6B MAR strain was originally
138 obtained from an HIV-negative child born to an HIV-positive mother, and has been
139 cultured in CBMCs (24). HHV-6B Z29 strain was originally isolated from an AIDS
140 patients from Zaire and obtained from the HHV-6 Foundation stock deposited at the NIH
141 AIDS repository and grown in the SupT1 cell line. Secondary HHV-6 standards
142 comprising quantitated viral DNA from HHV-6A GS strain (08-945-250) and HHV-6B
143 Z29 strain (08-923-00) were purchased from Advanced Biotechnologies Incorporated.
144 Strains sequenced in this study are available in Table 1.
145
146 *Illumina Sequencing Library Preparation*
147 DNA was extracted from culture isolates using the Zymo Viral DNA kit. DNA-
148 sequencing libraries were prepared from 50 ng of genomic DNA using quarter-volumes
149 of the Kapa HyperPrep kit with 7 minutes of fragmentation time and 12 cycles of dual-

indexed Truseq-adapter PCR (18). Libraries were sequenced on 2x300bp, 1x190bp, and/or 1x192bp runs on an Illumina MiSeq. Sequences were quality- and adapter-trimmed, de novo assembled, and contigs were aligned to reference HHV-6A (NC_001664) and HHV-6B (NC_000898) genomes and visualized using Geneious v9.1. Read mapping for copy number analysis was performed using the Geneious read mapper with 10% allowed gaps per read, word length of 18, 20% maximum mismatches per read, and with structural variant, insertion, and gap finding allowed.

qPCR confirmation

Quantitative PCR to estimate HHV-6 and beta-globin copy number in Table 1 was performed using UL67-directed directed 5R primers (GTT AGG ATA TAC CGA TGT GCG TGA T/ FAM- TCC GAA ACA ACT GTC TGA CTG GCA AAA-TAMRA/ TAC AGA TAC GGA GGC AAT AGA TTT G) (25) and beta-globin primers (TGA AGG CTC ATG GCA AGA AA/ FAM-TCC AGG TGA GCC AGG CCA TCA CTA-TAMRA/ GCT CAC TCA GTG TGG CAA AGG) respectively. Briefly each 30 ul PCR reaction contained 15 ul of 2x QuantiTect Multiplex PCR NoROX Master Mix (Qiagen), 830 nM each primer, 250 nM probe, 0.267 ul Rox (Invitrogen), and 0.03 units UNG (Epicentre). EXO internal control , including template, primers and probe, was spiked into each PCR reaction to monitor PCR inhibition. QuantStudio 7 Flex Real-Time PCR system was used to perform PCR and signal detection. The PCR thermocycling conditions are as following: 50^{o}C for 2 minutes, 95^{o}C for 15 minutes and followed by 45 cycles of 94^{o}C for 1 minute and 60^{o}C for 1 minute.

Quantitive PCR to estimate relative copy number between the origin of replication and HHV-6 U32 locus was performed in 20uL reactions using the SsoAdvanced Universal SYBR Green SuperMix. Ten-fold dilutions of DNA template from HHV-6 strains were tested using quantU32 F-R and species-specific origin primers (Table S1) using cycling conditions of 95C 30s and 40 cycles of 95C 5s and 60C 30s. For the U95 deletion in the CO strains, CO4 142970F-143228R primers (Table S1) were used in the same cycling conditions with SsoAdvanced Universal SYBR Green SuperMix, while the U32-targeting qPCR was performed with the U32 primers and a Taqman probe (Table S1) with the same cycling conditions.

PCR and Agilent Tapestation analysis

PCR across the origin tandem repeat was performed using 1 ng of template genomic DNA in 20uL total volume reactions using 10pmol of each primer and the Phusion High-Fidelity DNA polymerase according to manufacturer's instructions. PCR primer sequences can be found in Table S1. PCR reactions were analyzed with the Genomic DNA ScreenTape assay on an Agilent 4200 TapeStation.

Amplification-free nanopore sequencing

Nanopore libraries were created using the SQK-RAD002 kit tagmentation library preparation with 100ng of input total genomic DNA from the HHV-6B Z29 strain cultured in SupT1 cells. Amplification-free tagmented libraries were run according to Oxford Nanopore protocols v1.3.24 on a singular Mk1 (R9.4) FLO-MIN106 flow cell.

Nanopore reads were mapped to the HHV-6 Z29 reference genome (NC_000898) using the LASTZ and Geneious aligners to screen for origin-containing reads (26, 27).

Results

Large tandem repeats covering the origin of replication in both HHV-6A and HHV-6B strains

In order to obtain single nucleotide resolution and copy number measurement for HHV-6 type strain reference materials, we sequenced a HHV-6A GS strain obtained from the HHV-6 Foundation and a HHV-6B Z29 strain obtained from the NIH AIDS repository. Libraries from the HHV-6 Z29 and GS strain were each prepared twice and sequenced to an average depth of 76X and 453X, respectively. The HHV-6B Z29 strain contained a homogeneous 983 bp long tandem repeat (Figure 1A). Copy number estimates based on relative coverage at the edge of the repeats across multiple library preparations indicated an average of 11-13 copies of the repeat present. Mapping of the edges of the Z29 origin tandem repeat gave different repeat breakpoints than previously described (28). All Z29 strains sequenced in this study had an additional 123 nucleotides at the 5' end of the repeat and extra 4 nucleotides at the 3' end of the repeat than the previously described repeat to make the 983 bp tandem repeat (28).

HHV-6A GS strain contained a heterogeneous tandem repeat that covered 1260 bp of the HHV-6A reference genome (Figure 1B). Copy number estimates based on relative coverage at the edge of the repeats indicated an average of between 10-12 copies of the repeat present. The most common tandem repeat present included deletions of 193

bp and 2bp along with an insertion of 189 bp based on the HHV-6A reference genome, giving a mode length of 1254 bp.

To demonstrate that the copy number heterogeneity we found in the type strains of HHV-6 are also present in commercially available quantitative clinical reference materials, we also performed shotgun DNA sequencing on an HHV-6A GS strain and HHV-6B Z29 strain from Advanced Biotechnology Inc (ABI). These strains had similar tandem repeats in size at the origin of replication present in the GS strain obtained from the HHV-6 Foundation and Z29 strains obtained from NIH AIDS repository (Figure 1C). However, the Z29 origin tandem repeat in the commercial reference material was present at approximately twice the copy number observed in Z29 from the NIH AIDS repository. Intriguingly, the HHV-6A GS strain quantitative secondary standard material also contained a 4-fold increase in coverage covering the U90, U91, and the N-terminal two-thirds of the U86 gene. The U91 end of the repeat contained a complex rearrangement with the U100-DR intergenic region 266 nucleotides 5' of the beginning of the annotated DR repeat region. The HHV-6B Z29 secondary standard contained two large rearrangements -- one between U73 and R3 repeat region constituting 50% of DNA present and another between U86 and U95 representing 25% of DNA present. Thus, copy number of HHV-6B between U86 to R3 was 4-fold lower and between U73-U86 and R3-U95 was 2-fold lower than the rest of the genome. Quantitative PCR analysis of origin and U32 loci confirmed deep sequencing data, demonstrating a 3-4 cycle earlier Ct for origin tandem repeat containing strains compared to strains lacking origin tandem repeats. Of note, an early passage GS strain had 3-fold fewer copies of the origin than

238 the later passage GS-HSB2 strain, consistent with what has been described previously in
239 Z29 (28).

240 Previous analysis of 125 HHV-6B genomes obtained from clinical specimens
241 revealed no tandem repeats across the origin of replication. To confirm the origin tandem
242 repeat present in the HHV-6B Z29 strain, we performed PCR amplification and fragment
243 analysis by gel electrophoresis across the sequence tandem repeat with two separate
244 primer sets. As a control, we performed PCR across the origin of replication in a HHV-
245 6B PCR-positive patient specimen. The patient specimen demonstrated a single copy of
246 the locus present, while the Z29 strain contained an amplification ladder consistent with a
247 population of virus with different numbers of multiple tandem repeats present at the locus
248 (Figure 2A). We also performed amplification-free nanopore sequencing on DNA
249 extracted from the HHV-6B Z29 strain in culture. Across 6,369 nanopore reads, we
250 recovered two reads that contained more than one copy of the origin tandem repeat
251 (Figure 2B). One read contained three tandem repeats of the origin repeat, while the
252 other contained two repeats. Neither read spanned the entire tandem repeat.

253

254 *Interspecies recombination between HHV-6A and HHV-6B in strain DA*

255 Whole genome sequencing of the HHV6 DA strain revealed a hybrid genome
256 indicative of interspecies recombination between HHV-6A and HHV-6B strains. The
257 DA strain genome overall showed closest sequence identity to HHV-6A than HHV-6B
258 strains but included an oriLyt repeat that measured the exact length of the Z29 repeat at
259 983 bp (Figure 3A). The DA strain U38 gene matched with perfect identity to the HHV-
260 6A GS strain by BLASTN analysis. Of note, the DA strain origin-binding protein U73

gene matched closer to the HHV-6A U73 than HHV-6B U73 (99.5% versus 97.1% pairwise nucleotide identity to HHV-6A and HHV-6B reference strain genomes, respectively).

Analysis of the DA strain sequence yielded consensus recombination breakpoints at 67,314 bp and 72,667 bp of the NC_001664 that were detected by six of seven recombination analysis programs (RDP4, GENECONV, Bootscan, MaxChi, Chimaera, and 3Seq) (Figure 3B). Reads mapped with near identity along a 5.5kb region of the HHV-6B Z29 reference genome between the 5' end of U41 and 3' end of U43 (68535-73824 bp of NC_000898) (Figure 3C). Only a 61bp fragment with oriLyt repeat had 6 variant sites to HHV-6B sequences and matched identically to the HHV-6A sequences in this (Figure 3D). These sequences were just 3' from the end of the minimal origin of DNA replication annotated in the HHV-6A reference genome. The length of the HHV-6B sequence present in the DA strain is likely considerably longer than the 5.5kb shown on the reference genome, due to the presence of the oriLyt repeat. Strain DA also contained a 33bp deletion in the 3' end of U79 gene that is not represented in either HHV-6A or HHV-6B reference sequences.

Large deletions in U12-U24 genes and U94-U95 genes from two laboratory-adapted HHV-6A strains from collagen vascular disease

Five HHV-6A strains isolates from patients with collagen vascular disease also grew to high copy number in HSB2 cells. Isolates CO1, CO2, and CO3 were cultured in primary peripheral blood lymphocytes for 17-21 days and then HSB2 cells for 4-5 days, while isolates CO4 and CO7 were cultured for 2 days in primary cells and 42-46 days in

284 HSB2 cells (22). The five CO isolates were highly similar with an average pairwise
285 nucleotide identity of >99.9%, while all five isolates most closely aligned to HHV-6A
286 isolate GS (KJ123690.1, 99.2% pairwise nucleotide identity). Both isolates CO4 and
287 CO7 had 60% decreased coverage in a 13.5kb region from U12-U24 relative to the CO1-
288 CO3 isolates (Figure 4A). Most notably, both isolates CO4 and CO7 also had 95% lower
289 coverage over a 4.9kb region covering genes U94 and U95 relative to CO1-CO3 isolates.
290 Equivalent relative copy number estimates for both the U95 locus and the origin of
291 replication were recovered by qPCR (Figure 4B/C). Isolates CO4 and CO7 both had
292 equivalent mixed variant allele frequencies at 25 loci based on read mapping to the UL
293 region of the HHV-6A reference genome (NC_001664) (Table S2). No other variants
294 were isolated between the HHV-6 CO4 and CO7 strains, suggesting these strains are
295 identical.
296
297 *HHV-6 strains isolated from PHA-stimulated cord blood mononuclear cells do not*
298 *contain large tandem duplications or deletions*
299 To find the best secondary standard for HHV-6 clinical testing and to better
300 understand the origin of the origin tandem repeat, we sequenced four HHV-6B and one
301 HHV-6A strain that were isolated from PHA-stimulated cord blood mononuclear cells.
302 Four HHV-6B strains (HST, ENO, KYO, NAK) were isolated from Japanese exanthem
303 subitum patients, one HHV-6B strain (MAR) was isolated from an asymptomatic French
304 child, and one HHV-6A strain (SIE) was isolated from an Ivory Coast patient with adult
305 T-cell leukemia (23, 24, 29). Interestingly, all six strains contained minimal levels of
306 copy number heterogeneity with an average coefficient of variation of coverage in the

307 unique long region of 21.0%, compared with 119% averaged over HHV-6A GS and
308 HHV-6B Z29 type strains (Figure 5A). The lack of origin amplification in these CBMC-
309 passaged strains was also confirmed by qPCR (Figure 5B). Without the tandem repeat in
310 the origin, the only source of copy number differences in the cord blood mononuclear
311 passaged strains was direct repeat coverage at two-thirds of that in the unique long
312 region, consistent with active replication and sequencing of mostly HHV-6 concatemers
313 (Figure 5A) (30). Of note, the decreased coverage in the direct repeat region relative to
314 the unique long region was present in all strains sequenced in this study.
315
316 **Discussion**
317 We show the presence of copy number heterogeneity at multiple genomic loci
318 across HHV-6A and HHV-6B culture isolates that are used as reference material for
319 clinical assay development and normalization as well as basic science virology work.
320 The most prominent copy number difference was a large tandem repeat in the origin of
321 replication that was present in both HHV-6A and HHV- 6B strains. The presence of
322 large copy number differences in HHV-6A and HHV-6B strains was associated with
323 passaged in immortalized cell lines, as strains passaged in primary cell lines and cord
324 blood mononuclear cells did not carry genomic duplications and deletions and early
325 passage virus contained fewer tandem repeats than late passage virus.
326 Recent analysis of 130 HHV-6B genomes and 10 iciHHV-6A sequences spanning
327 the U41 origin region revealed no large tandem repeats in the origin of replication among
328 clinical isolates (18). Previous work had illustrated a large heterogeneous tandem repeat
329 present at the oriLyt in HHV-6B Z29 strains that was associated with higher passage

number (28). Our data illustrate that this is a general feature of HHV-6 and that a heterogeneous larger tandem repeat is also present in multiple laboratory adapted HHV-6A strains at the oriLyt. We show that increased culture passage in immortalized cell lines is associated with reduced copy number of two large genomic loci in HHV-6A CO strains. Copy number variability at multiple genomic loci was also reflected in HHV-6A reference material that is used to normalize quantitative values for clinical assay development. Interestingly, no HHV-6 isolate from cord blood mononuclear cells showed tandem repeats, despite growing to high titer. These isolates may provide the best material for HHV-6 standards development.

We also show the first genomic evidence of interspecies recombination between HHV-6A and HHV-6B strains along a >5.5 kb segment containing the oriLyt and U42 gene as well as portions of the U41 and U43 genes. Interspecies recombination is a relatively common feature of the alphaherpesviruses HSV-1 and HSV-2 but has not been described for any other human herpesviruses (31). Our results for DA strain recombination are most consistent with a model in which co-cultivation of an HHV-6A GS strain with a laboratory-adapted HHV-6B Z29 strain with oriLyt repeat that resulted in recombination between the two strains. Interestingly the recombinant strain also showed HHV-6A-like sequence in its HHV-6B oriLyt tandem repeat. Although this sequence fell outside the minimal origin of replication, it is suggestive that the oriLyt tandem repeat sequence may have evolved to more effectively interact with HHV-6A replication proteins contained in the rest of genome. No specific HHV-6B-like sequences were found in the DA strain U73 origin binding protein to indicate reciprocal U73 evolution to match the HHV-6B-like origin sequences. Recent analysis of 140 HHV-6

353 genomes from clinical and iciHHV-6 isolates revealed no evidence of interspecies
354 recombination but widespread intraspecies recombination (18). Whether interspecies
355 recombination has occurred in clinical strains of HHV-6 remains to be determined.
356 However the high sequence similarity between HHV-6A and HHV-6B and the frequency
357 of alphaherpesvirus recombination suggests that as more sequencing is performed, HHV-
358 6A and HHV-6B recombinants may be found in nature.

359 Previous work from our group had demonstrated the loss of almost one-third of
360 the BK and JC polyomavirus genome in up to 90% of viral species present in multiple
361 viral stocks including a WHO international standard, likely due to viral passage in SV40
362 T-antigen immortalized cell lines (19, 20). While clinical PCR tests for HHV-6 are
363 unlikely to target the oriLyt region, normalization of quantitative clinical HHV-6 testing
364 to any of the loci here found at increased or decreased copy number could affect
365 quantitation, not least because multiple viral populations were present in many of the
366 reference materials tested. We also found several examples where contamination with
367 another HHV-6 strain most parsimoniously explained our sequencing results. We
368 continue to recommend the use of next-generation sequencing to obtain genome-wide
369 single nucleotide resolution copy number measurements in order to validate viral
370 reference materials used in clinical virology and basic sciences labs across the world.
371
372 Accessions: These sequences are available in NCBI Genbank (MF994813-MF994829)
373 and associated with BioProject 338014.
374

375 **Figure Legends**

376 **Figure 1** – Representative coverage maps of HHV-6B Z29 and HHV-6A GS reference
377 strains. Shotgun DNA sequencing reads from cultured virus were mapped to the NCBI
378 HHV-6 reference genomes, NC_000898 and NC_001664, respectively. A) HHV-6B
379 Z29 strain yielded a homogeneous 983 bp tandem repeat that was present at
380 approximately 12.5X higher coverage of the rest of genome. Sequences at 5' and 3' end
381 of the tandem repeat in Z29 strain are depicted and are different than those indicted in
382 original paper (28). B) HHV-6A GS strain yielded a heterogeneous 1254 bp tandem
383 repeat that was present at approximately 11.4X higher coverage than the rest of the
384 genome. Sequences at 5' and 3' end of the heterogeneous tandem repeat in GS strain are
385 depicted. C) ABI Quantitative DNA material for HHV-6A GS and HHV-6B Z29 also
386 demonstrates a similar origin tandem repeat with additional loci with copy number
387 differences in the GS strain. Long-distance rearrangements between U73-R3, U86-U95,
388 and U91-U100/DR intergenic region are noted by arching dotted lines and the estimated
389 viral subpopulation containing the indicated deletion is indicated by the % value. D)
390 qPCR analysis of U32 and origin loci on 10-fold dilutions of DNA from HHV-6A and 6B
391 strains confirms deep sequencing estimates of relative copy number. Equivalent Cts for
392 the PCR results of HST/MT4 indicate an equal amount of amplification from the 2 sites
393 with that strain. All other strains show increased amplification with the Origin PCR,
394 indicating additional copies of the origin present in the genome. Of note, an early
395 passage GS strain showed 3-fold less amplification of the origin by qPCR compared to
396 the later passage GS-HSB2 strain.
397

Figure 2 – Validation of Z29 origin tandem repeat with PCR and nanopore sequencing. A) PCR-TapeStation analysis of tandem repeat in origin of Z29 strain with two PCR primer pairs. B) Amplification free-nanopore sequencing yielded two nanopore reads that align across the origin tandem repeat and carry at least 2 and 3 copies of the repeat. No reads that spanned both ends of the tandem repeat were recovered.

Figure 3 – HHV-6A DA strain shows genomic evidence of interspecies recombination between HHV-6A and HHV-6B strains. A) Tandem repeat of the origin for DA strain shows Z29-like length of 983 bp. B) RDP4 scan recombination analysis demonstrates two breakpoints recombination breakpoints at nucleotide 67,194 and 72,681 of the HHV-6A DA genome C) Mapping of reads to HHV-6B Z29 reference genome at U41 locus with mismatches highlighted. D) Nucleotide sequence of 61bp region of tandem repeat that most closely matches HHV-6A reference genome (NC_001664). E) Phylogenetic tree analysis of recombination region supports HHV-6B-like nature of U41 locus.

Figure 4 – HHV-6A CO strains from patients with collagen vascular diseases show several copy number differences. A) Coverage maps from five HHV-6A strains isolated from different patients with collagen vascular diseases are displayed (22). These strains all showed a similar heterogeneous tandem repeat that gave a mode length of 1254 bp. Strains CO4 and CO7 which were passaged >40 times in immortalized HSB2 cell lines also demonstrated 60% and 95% lower coverage in U12-U24 and U94-U95 regions, respectively. Of note, these two strains also shared identical sequence and minor allele distribution, consistent with being the same strain. B/C) qPCR analysis of ten-fold

421 dilutions of DNA from the HHV-6A CO strains at the U32, U95, and origin loci confirms
422 relative copy number estimates from deep sequencing data.
423
424 **Figure 5 –** HHV-6A and -6B strains sequenced from PHA-activated cord blood
425 mononuclear cells reveal no major copy number differences. Coverage maps from four
426 HHV-6B strains from Japan (NAK, KYO, ENO, HST), one HHV-6B strain from France
427 (MAR), and one HHV-6A strain from Ivory Coast (SIE) are displayed. The only copy
428 number difference in these strains is the reduced coverage in the direct repeat region due
429 to sequencing of likely concatemeric HHV-6, which was present in all strain sequenced
430 in this study. B) qPCR analysis of ten-fold dilutions of DNA from the CBMC passaged
431 HHV-6 strains at the U32 and origin loci confirms relative copy number estimates from
432 deep sequencing data.
433
434
435 **Table 1** – Summary of HHV-6 strains sequenced in this study.

436 **References**

437 1. Salahuddin SZ, Ablashi DV, Markham PD, Josephs SF, Sturzenegger S, Kaplan M,
438 Halligan G, Biberfeld P, Wong-Staal F, Kramarsky B. 1986. Isolation of a new virus,
439 HBLV, in patients with lymphoproliferative disorders. Science 234:596–601.

440 2. Ablashi D, Agut H, Alvarez-Lafuente R, Clark DA, Dewhurst S, DiLuca D, Flamand
441 L, Frenkel N, Gallo R, Gompels UA, Höllsberg P, Jacobson S, Luppi M, Lusso P,
442 Malnati M, Medveczky P, Mori Y, Pellett PE, Pritchett JC, Yamanishi K, Yoshikawa
443 T. 2014. Classification of HHV-6A and HHV-6B as distinct viruses. Arch Virol
444 159:863–870.

445 3. Fillet AM, Raphael M, Visse B, Audouin J, Poirel L, Agut H. 1995. Controlled study
446 of human herpesvirus 6 detection in acquired immunodeficiency syndrome-
447 associated non-Hodgkin's lymphoma. The French Study Group for HIV-Associated
448 Tumors. J Med Virol 45:106–112.

449 4. Cone RW, Huang ML, Hackman RC, Corey L. 1996. Coinfection with human
450 herpesvirus 6 variants A and B in lung tissue. J Clin Microbiol 34:877–881.

451 5. Boutolleau D, Fernandez C, André E, Imbert-Marcille B-M, Milpied N, Agut H,
452 Gautheret-Dejean A. 2003. Human herpesvirus (HHV)-6 and HHV-7: two closely
453 related viruses with different infection profiles in stem cell transplantation recipients.
454 J Infect Dis 187:179–186.

455 6. Hill JA, Magaret AS, Hall-Sedlak R, Mikhaylova A, Huang M-L, Sandmaier BM,
456 Hansen JA, Jerome KR, Zerr DM, Boeckh M. 2017. Outcomes of hematopoietic cell

457 transplantation using donors or recipients with inherited chromosomally integrated
458 HHV-6. Blood 130:1062–1069.

459 7. Ogata M, Kikuchi H, Satou T, Kawano R, Ikewaki J, Kohno K, Kashima K, Ohtsuka
460 E, Kadota J-I. 2006. Human herpesvirus 6 DNA in plasma after allogeneic stem cell
461 transplantation: incidence and clinical significance. J Infect Dis 193:68–79.

462 8. Flamand L, Gravel A, Boutolleau D, Alvarez-Lafuente R, Jacobson S, Malnati MS,
463 Kohn D, Tang Y-W, Yoshikawa T, Ablashi D. 2008. Multicenter comparison of PCR
464 assays for detection of human herpesvirus 6 DNA in serum. J Clin Microbiol
465 46:2700–2706.

466 9. Karlsson T, Mannonen L, Loginov R, Lappalainen M, Höckerstedt K, Lautenschlager
467 I. 2012. Development of a new quantitative real-time HHV-6-PCR and monitoring of
468 HHV-6 DNAaemia after liver transplantation. J Virol Methods 181:25–36.

469 10. Hayden RT, Sun Y, Tang L, Procop GW, Hillyard DR, Pinsky BA, Young SA,
470 Caliendo AM. 2017. Progress in Quantitative Viral Load Testing: Variability and
471 Impact of the WHO Quantitative International Standards. J Clin Microbiol 55:423–
472 430.

473 11. de Pagter PJ, Schuurman R, de Vos NM, Mackay W, van Loon AM. 2010.
474 Multicenter external quality assessment of molecular methods for detection of human
475 herpesvirus 6. J Clin Microbiol 48:2536–2540.

476 12. Hayden RT, Shahbazian MD, Valsamakis A, Boonyaratanakornkit J, Cook L, Pang
477 XL, Preiksaitis JK, Schönbrunner ER, Caliendo AM. 2013. Multicenter evaluation of

478 a commercial cytomegalovirus quantitative standard: effects of commutability on
479 interlaboratory concordance. J Clin Microbiol 51:3811–3817.

480 13. Hayden RT, Yan X, Wick MT, Rodriguez AB, Xiong X, Ginocchio CC, Mitchell MJ,
481 Caliendo AM, College of American Pathologists Microbiology Resource Committee.
482 2012. Factors contributing to variability of quantitative viral PCR results in
483 proficiency testing samples: a multivariate analysis. J Clin Microbiol 50:337–345.

484 14. Hayden RT, Preiksaitis J, Tong Y, Pang X, Sun Y, Tang L, Cook L, Pounds S, Fryer
485 J, Caliendo AM. 2015. Commutability of the First World Health Organization
486 International Standard for Human Cytomegalovirus. J Clin Microbiol 53:3325–3333.

487 15. Madej RM, Davis J, Holden MJ, Kwang S, Labourier E, Schneider GJ. 2010.
488 International standards and reference materials for quantitative molecular infectious
489 disease testing. J Mol Diagn JMD 12:133–143.

490 16. Holden MJ, Madej RM, Minor P, Kalman LV. 2011. Molecular diagnostics:
491 harmonization through reference materials, documentary standards and proficiency
492 testing. Expert Rev Mol Diagn 11:741–755.

493 17. Zhang E, Bell AJ, Wilkie GS, Suárez NM, Batini C, Veal CD, Armendáriz-Castillo I,
494 Neumann R, Cotton VE, Huang Y, Porteous DJ, Jarrett RF, Davison AJ, Royle NJ.
495 2017. Inherited chromosomally integrated human herpesvirus 6 genomes are ancient,
496 intact and potentially able to reactivate from telomeres. J Virol JVI.01137-17.

497 18. Greninger A, Knudsen G, Roychoudhury P, Hanson D, Sedlak RAH, Xie H, Guan J,
498 Nguyen T, Peddu V, Boeckh M, Huang M-L, Cook L, Depledge D, Zerr D, Koelle D,

499 Gantt S, Yoshikawa T, Caserta M, Hill J, Jerome K. 2017. Genomic and proteomic
500 analysis of Human herpesvirus 6 reveals distinct clustering of acute versus inherited
501 forms and reannotation of reference strain. bioRxiv 181248.

502 19. Greninger AL, Bateman AC, Atienza EE, Wendt S, Makhsous N, Jerome KR, Cook
503 L. 2017. Copy Number Heterogeneity of JC Virus Standards. J Clin Microbiol
504 55:824–831.

505 20. Bateman AC, Greninger A, Atienza EE, Jerome KR, Cook L. 2017. Quantification of
506 BK Virus Standards by Quantitative Real-time PCR and Droplet Digital PCR is
507 Confounded by Multiple Virus Populations in the WHO BKV International Standard.
508 Clin Chem Press.

509 21. Gravel A, Ablashi D, Flamand L. 2013. Complete Genome Sequence of Early
510 Passaged Human Herpesvirus 6A (GS Strain) Isolated from North America. Genome
511 Announc 1.

512 22. Krueger GR, Sander C, Hoffmann A, Barth A, Koch B, Braun M. 1991. Isolation of
513 human herpesvirus-6 (HHV-6) from patients with collagen vascular diseases. Vivo
514 Athens Greece 5:217–225.

515 23. Yamanishi K, Okuno T, Shiraki K, Takahashi M, Kondo T, Asano Y, Kurata T.
516 1988. Identification of human herpesvirus-6 as a causal agent for exanthem subitum.
517 Lancet Lond Engl 1:1065–1067.

518 24. Aubin JT, Collandre H, Candotti D, Ingrand D, Rouzioux C, Burgard M, Richard S,
519 Huraux JM, Agut H. 1991. Several groups among human herpesvirus 6 strains can be

520 distinguished by Southern blotting and polymerase chain reaction. J Clin Microbiol
521 29:367–372.

522 25. Zerr DM, Yeung LC, Obrigewitch RM, Huang M-L, Frenkel LM, Corey L. 2002.
523 Primary human herpesvirus-6 associated with an afebrile seizure in a 3-week-old
524 infant. J Med Virol 66:384–387.

525 26. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D,
526 Bouquet J, Somasekar S, Linnen JM, Dodd R, Mulembakani P, Schneider BS,
527 Muyembe-Tamfum J-J, Stramer SL, Chiu CY. 2015. Rapid metagenomic
528 identification of viral pathogens in clinical samples by real-time nanopore sequencing
529 analysis. Genome Med 7:99.

530 27. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S,
531 Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.
532 2012. Geneious Basic: an integrated and extendable desktop software platform for the
533 organization and analysis of sequence data. Bioinforma Oxf Engl 28:1647–1649.

534 28. Stamey FR, Dominguez G, Black JB, Dambaugh TR, Pellett PE. 1995. Intragenomic
535 linear amplification of human herpesvirus 6B oriLyt suggests acquisition of oriLyt by
536 transposition. J Virol 69:589–596.

537 29. Agut H, Guetard D, Collandre H, Dauguet C, Montagnier L, Miclea JM, Baurmann
538 H, Gessain A. 1988. Concomitant infection by human herpesvirus 6, HTLV-I, and
539 HIV-2. Lancet Lond Engl 1:712.

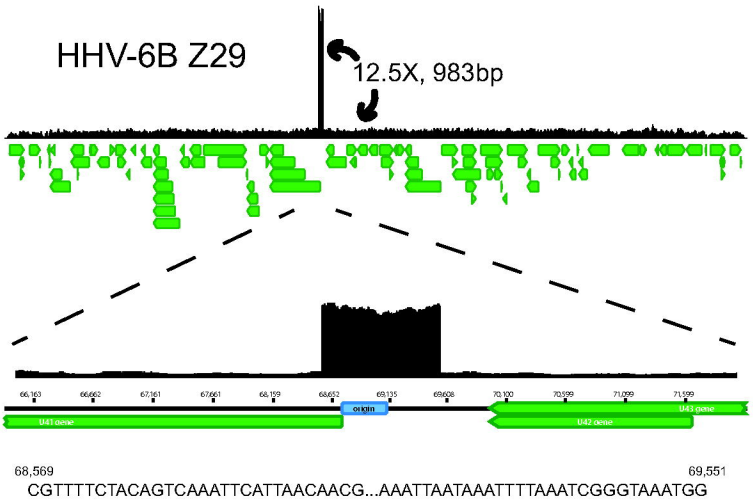
30. Prusty BK, Krohne G, Rudel T. 2013. Reactivation of Chromosomally Integrated Human Herpesvirus-6 by Telomeric Circle Formation. PLOS Genet 9:e1004033.

31. Koelle DM, Norberg P, Fitzgibbon MP, Russell RM, Greninger AL, Huang M-L, Stensland L, Jing L, Magaret AS, Diem K, Selke S, Xie H, Celum C, Lingappa JR, Jerome KR, Wald A, Johnston C. 2017. Worldwide circulation of HSV-2□×□HSV-1 recombinant strains. Sci Rep 7:44084.

Table 1. Strains sequenced with qPCR results and Sequencing Reads

Species	Strain	Cell Line	HHV-6 quantity	beta-globin quantity	Trimmed Reads	HHV-6 Reads	Accession
HHV-6A							
	GS	HSB2	1.94E+06	5.38E+03	2091792	47508	MF994822
	GS-early	CBMC	6.57E+01	1.36E+03	308104	66878	KC465951 (sequenced in PMID 23766398, received reads from L. Flamand)
	DA	HSB2	9.17E+06	5.39E+04	3787217	30508	MF994820
	CO-1	HSB2	3.04E+07	1.51E+04	1913207	164835	MF994815
	CO-2	HSB2	2.07E+07	8.04E+03	1278582	118403	MF994816
	CO-3	HSB2	4.20E+06	3.25E+03	896950	54464	MF994817
	CO-4	HSB2	2.47E+06	1.70E+03	6692042	42596	MF994818
	CO-7	HSB2	1.61E+06	1.04E+03	10887157	54046	MF994819
	SIE	PHA-stimulated CBMC	2.05E+07	2.19E+04	1015200	49307	MF994828
	ABI-HHV6A (GS)	unknown	2.26E+04	2.18E+01	1644798	225145	MF994813
HHV-6B							
	Z29	SupT1	1.89E+05	1.54E+03	2436488	62770	MF994829
	HST	MT4	4.10E+05	5.70E+03	10666502	28251	MF994824
	HST	PHA-stimulated CBMC	1.12E+06	2.82E+03	1155336	29365	MF994823
	KYO	PHA-stimulated CBMC	1.56E+06	3.28E+03	949314	28626	MF994825
	ENO	PHA-stimulated CBMC	1.74E+06	7.00E+03	1116480	39004	MF994821
	MAR	PHA-stimulated CBMC	1.18E+07	1.97E+04	801650	26594	MF994826
	NAK	PHA-stimulated CBMC	1.29E+05	6.88E+02	5617132	42955	MF994827
	ABI-HHV6B (Z29)	unknown	1.16E+03	Undetectable	1200052	112632	MF994814

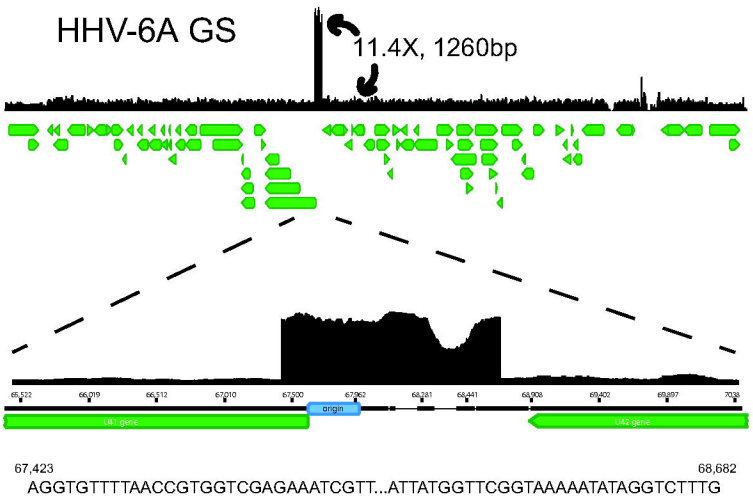
A



68,569 69,551

CGTTTTCTACAGTCAAATTCATTAACAACG...AAATTAATAAATTTTAAATCGGGTAAATGG

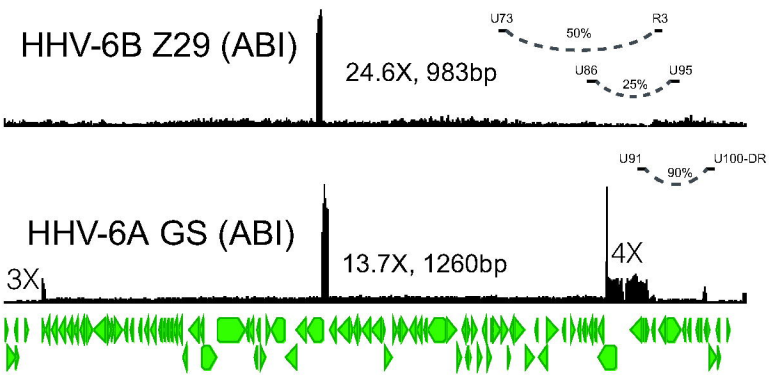
B



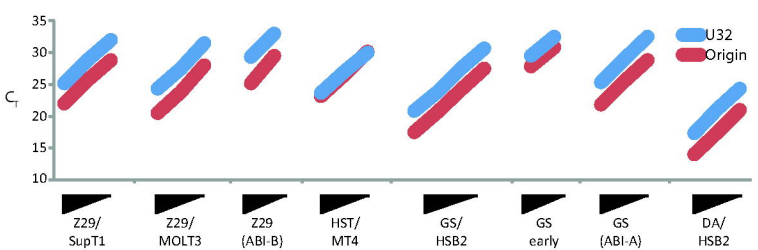
67,423 68,682

AGGTGTTTTAACCGTGGTCGAGAAATCGTT...ATTATGGTTCGGTAAAAATATAGGTCTTTG

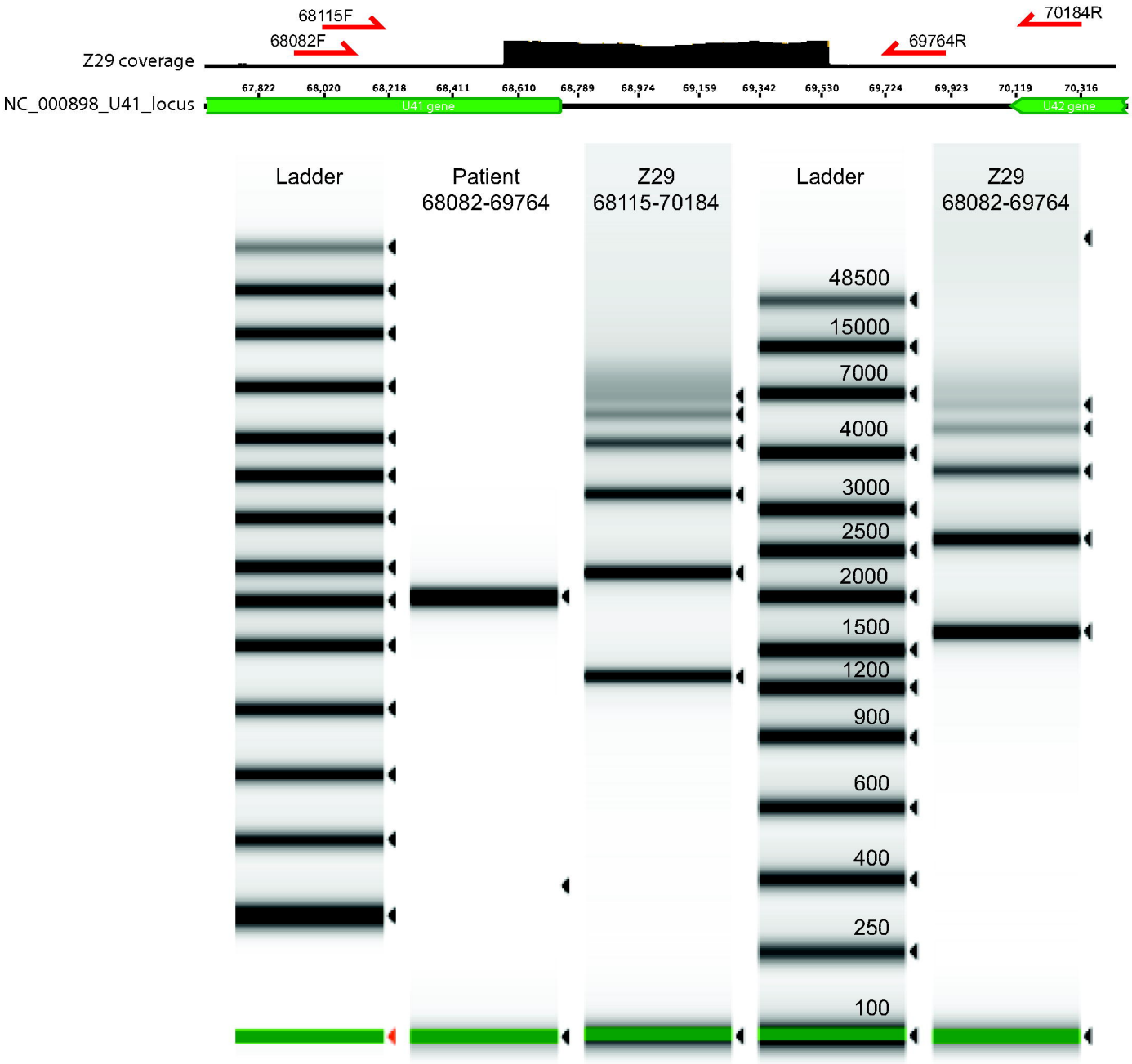
C



D



A



B