



Higher-Order Chromatin Structures of Chromosomally Integrated HHV-6A Predict Integration Sites

Michael Mariani¹, Cosima Zimmerman², Princess Rodriguez¹, Ellie Hasenohr¹, Giulia Aimola², Diana Lea Gerrard¹, Alyssa Richman¹, Andrea Dest¹, Louis Flamand³, Benedikt Kaufer^{2*} and Seth Fietze^{1,4*}

¹ Department of Biomedical and Health Sciences, College of Nursing and Health Sciences, University of Vermont, Burlington, VT, United States, ² Institute of Virology, Freie Universität Berlin, Berlin, Germany, ³ Division of Infectious Disease and Immunity, CHU de Québec Research Center-Université Laval, Québec City, QC, Canada, ⁴ University of Vermont Cancer Center, Burlington, VT, United States

OPEN ACCESS

Edited by:

Georges Herbein,
University of Franche-Comté, France

Reviewed by:

Marc Lavigne,
Institut Pasteur, France
Cyprian Rossetto,
University of Nevada, Reno,
United States

*Correspondence:

Seth Fietze
seth.fietze@med.uvm.edu
Benedikt Kaufer
benedikt.kaufer@fu-berlin.de

Specialty section:

This article was submitted to
Virus and Host,
a section of the journal
Frontiers in Cellular and
Infection Microbiology

Received: 30 September 2020

Accepted: 20 January 2021

Published: 26 February 2021

Citation:

Mariani M, Zimmerman C,
Rodriguez P, Hasenohr E,
Aimola G, Gerrard DL, Richman A,
Dest A, Flamand L, Kaufer B and
Fietze S (2021) Higher-Order
Chromatin Structures of
Chromosomally Integrated
HHV-6A Predict Integration Sites.
Front. Cell. Infect. Microbiol. 11:612656.
doi: 10.3389/fcimb.2021.612656

Human herpesvirus -6A and 6B (HHV-6A/B) can integrate their genomes into the telomeres of human chromosomes. Viral integration can occur in several cell types, including germinal cells, resulting in individuals that harbor the viral genome in every cell of their body. The integrated genome is efficiently silenced but can sporadically reactivate resulting in various clinical symptoms. To date, the integration mechanism and the subsequent silencing of HHV-6A/B genes remains poorly understood. Here we investigate the genome-wide chromatin contacts of the integrated HHV-6A in latently-infected cells. We show that HHV-6A becomes transcriptionally silent upon infection of these cells over the course of seven days. In addition, we established an HHV-6-specific 4C-seq approach, revealing that the HHV-6A 3D interactome is associated with quiescent chromatin states in cells harboring integrated virus. Furthermore, we observed that the majority of virus chromatin interactions occur toward the distal ends of specific human chromosomes. Exploiting this finding, we established a 4C-seq method that accurately detects the chromosomal integration sites. We further implement long-read minION sequencing in the 4C-seq assay and developed a method to identify HHV-6A/B integration sites in clinical samples.

Keywords: epigenetics, chromatin 3D architecture, latency, herpesvirus (hhv-6), gene expression

INTRODUCTION

Human herpesvirus 6A (HHV-6A) and 6B (HHV-6B) are two closely related and ubiquitous betaherpesviruses (Zerr et al., 2005). Most people are infected with HHV-6B as infants, while the etiology of HHV-6A remain poorly defined (Okuno et al., 1989; De Bolle et al., 2005). Following primary infection, HHV-6A/B establishes latency in the host for life (Flamand, 2018). During this latent state, viral gene expression and viral load are not detectable (Saviola et al., 2019). However, the virus can sporadically reactivate in the host, resulting in the production of infectious virions and viral transmission (Endo et al., 2014). HHV-6A/B reactivation has been linked to a variety of pathologies including encephalitis, graft rejection and a spectrum of other diseases (Aimola et al., 2020).

HHV-6A/B are unique among human herpesviruses as they are able to integrate their genomes into the telomeres of human chromosomes (Arbuckle et al., 2010; Osterrieder et al., 2014). Chromosomal integration of HHV-6A/B occurs very efficient *in vitro* and is dependent on telomere sequence arrays present at the ends of the virus genome (Wallaschek et al., 2016; Gravel et al., 2017). HHV-6A/B can also integrate into the germline, resulting in individuals harboring the virus in every nucleated cell of the body and inherit it to their offspring. This inherited chromosomally-integrated HHV-6 (iciHHV-6) is present in about 1% of the human population (Kaufer and Flamand, 2014) with several independent integration events occurring thousands of years ago (Aswad et al., 2020). Integration into the telomeres allows maintenance of the virus genome in iciHHV-6 individuals and latency (Arbuckle et al., 2010); however, it remains unknown if integration into the telomeres contributes to the silencing of the virus genome.

Here, we explored if, how and when the HHV-6A genome is silenced upon integration using an *in vitro* integration model and RNA-sequencing. We demonstrate that HHV-6A genes are silenced upon integration over the course of 7 days. We employ circular chromosome conformation capture assays (4C-seq) to assess the higher order chromatin structures formed between the virus and host genome in these cells. Because HHV-6A/B integrate into highly repetitive telomeric regions of individual human chromosomes, we used 4C-seq-based analysis as a tool to identify integration sites. Our 4C-seq approach employs novel scoring method as well as Oxford Nanopore minION long-read sequencing to effectively identify integration sites in human telomeres. Overall, this study provides a better understanding of the chromatin programs that may regulate HHV-6A latency and provide novel diagnostic methods to determine chromosomal integration sites.

MATERIALS AND METHODS

Ethics

Specimens were obtained from the Fred Hutch Research Cell Bank, which prospectively collects and cryopreserves peripheral blood mononuclear cells (PBMCs) from donors and recipients. The University of Washington Institutional Review Board approved use of the iciHHV-6 specimens from the Fred Hutchinson Cancer Research Center and use of anonymized excess HHV-6-positive samples submitted for testing at the University of Washington Virology lab.

Cell Lines and Virus

SMC cells harboring integrated HHV-6A virus (iciHHV-6A cells) were immortalized following transduction with a lentiviral vector (pLenti CMV/TO SV40 small + large T (W612-1), a gift from Eric Campeau and obtained from Addgene (plasmid #22298)) and expressing the SV40 T antigens. SMCs were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 20% fetal bovine serum (FBS), 1% penicillin streptomycin (Pen/Strep), and 1%

L-Glutamine. Human epithelial kidney 293 T (293 T, ATCC CRL-11268) cells were cultured in the same medium but supplemented with 10% FBS. All cells were maintained in 10 cm² flasks as a monolayer culture in a humidified 5% CO₂ air incubator at 37°C. Bacterial artificial chromosome (BAC)-derived HHV6A (strain U1102) expressing green fluorescent protein (GFP) under the control of the HCMV major immediate early (IE) promoter (293-HHV-6A virus) was propagated in JJHan cells as described previously (Tang et al., 2010b). 293 T cells were infected with 293-HHV-6A virus and GFP positive cells were isolated using a FACS AriaIII cell sorter (BD Biosciences). Clones harboring the integrated 293-HHV-6A genome (ciHHV-6A cells) were identified by quantitative PCR (qPCR) and HHV-6A integration was confirmed by fluorescent *in situ* hybridization (FISH). iciHHV-6B+ human lymphocytes (NCBI GenBank: KY315552) were isolated from an infected patient as described by Aswad et al. (Aimola et al., 2020; Aswad et al., 2020).

Fluorescence *In Situ* Hybridization

FISH was performed with digoxigenin-labeled bacterial artificial chromosome (BAC) probes against both the HHV-6A genome and individual chromosomal reads essentially as described previously (Kaufer et al., 2011; Kaufer, 2013). Slides were mounted using DAPI Vectashield (Vector Laboratories) and images were taken with an Axio Imager M1 (Zeiss).

RNA-Seq Analysis

RNA was extracted from an HHV-6A infected 293T cell line using Trizol Reagent (Life Technologies) and purified using Direct-zol RNA MicroPrep Kit (Zymo Research #R2060) following the manufacturer's instructions. One microgram of total RNA was depleted of ribosomal RNA (rRNA) using the KAPA RiboErase Kit (#KR1142) and libraries were prepared using the KAPA Stranded RNA-seq Library Preparation Kit (#KR0934). Libraries were quantified using Qubit (Life Technologies), and quality was assessed using the Agilent Bioanalyzer High-Sensitivity DNA kit (Agilent Technologies). Barcoded libraries were pooled and sequenced on an Illumina HiSeq 2000 to obtain 100-bp paired-end reads. RNA-seq data were processed using STAR as described previously (Dobin et al., 2013) using human (hg38) and a custom HHV-6A reference genomes (NC_001664.4) modified to contain GFP transgenes. BAM files were sorted and converted to SAM files using SAMtools (Li et al., 2009) and reads were counted with featureCounts (Liao et al., 2014) against the corresponding viral GTF file and normalized in DESeq2 (Love et al., 2014). Alignment of viral reads across the HHV-6A genome was visualized using the "seqsetvis" R library (Boyd, 2020) and in-house scripts.

4C-Seq Library Preparation and Sequencing

In order to identify the higher order chromatin structure and virus-host DNA contacts, 4C-seq libraries were prepared using both ciHHV-6A and iciHHV-6A samples as described previously (Krijger et al., 2020), using HindIII/DpnII restriction enzymes.

Cells were cultured as previously described (Saviola et al., 2019) and fixed with 1% formaldehyde prior to being snap frozen. For each 4C library, ten million cells were used unless otherwise noted. Inverse PCR was performed in 50 μ l reactions using 200 ng 3C library template, 25 μ l Q5 Hot Start High-Fidelity DNA Polymerase 2X Mastermix (New England Biolabs), 1.5 μ l of 10uM forward and reverse inverse PCR primers and water under the following conditions: 30 s 98°C, 10 s 98°C, 30 s 55°C, 2 m 72°C, 5 m 72°C for a total of 25 cycles. PCR products were cleaned up using the PCR purification kit (ThermoFisher Scientific) using the B3 reagent to exclude fragments less than 300 base pairs or with Ampure XP beads (Beckman Coulter) and eluted in water. Library indexing was performed using the DNA HT Dual Index Kit (Illumina) or with custom dual indexing primers with an additional eight cycles of PCR. Indexing of nanopore libraries was performed with custom 12-nucleotide dual index primers with eight additional PCR cycles. All primers sequences are listed in **Supplementary Table 1**. Libraries were pooled and sequenced with either HiSeq2000 or MinION platforms. The experiments were performed with samples from separate cell culture dates. For the Illumina viewpoint comparison libraries, two replicates were performed for two separate viewpoints. For the ONP titration and viewpoint comparison studies, one and two replicates were performed, respectively.

4C-Seq Primer Design

We developed a 4C-seq primer design tool¹ to facilitate the generation of the inverse PCR primers necessary for 4C-Seq assays in a genome-agnostic manner for use with experiments involving custom genomes (i.e., chromosomally integrated HHV-6A). A fasta file containing the HHV-6A genome (NC_001664) was obtained from NCBI and uploaded into the design tool, selecting HindIII and DpnII restriction endonucleases, and minimum fragment sizes 500 bp for the first fragment and 300 bp for the second fragment were used (van de Werken et al., 2012).

4C-Seq Analysis

4C-seq *trans* interactions were determined using a “window” method as described by Kim et al., (2020). Briefly, the virus viewpoint regions at the beginning of the reads were trimmed off using Cutadapt (Martin, 2011) software and the remaining human portions were aligned to the human UCSC hg38 reference genome using bowtie2 (version 2.3.5) (Langmead et al., 2009; Langmead and Salzberg, 2012). P-values for each mapped base were then calculated for each 10kb window across all chromosomes using a Poisson formula. Final, significant peak regions are then called using the MACS2 (version 2.2.7) *bdgpeakcall* function (Zhang et al., 2008). Significant Interacting *trans* peak regions were identified using the above window method for each replicate, for each viewpoint, in both the ciHHV-6A and iciHHV-6A samples. These regions were plotted as circos plots using the Circlize R library (Gu et al., 2014). Bedtools (Quinlan and Hall, 2010) was used to find the intersection of the *trans* regions and the 127 epigenomes 25-state

imputation based chromatin state model created with ChromImpute (Ernst and Kellis, 2012; Ernst and Kellis, 2015) and made available from the Wang Lab² (Jin Wook Lee). For ciHHV-6A cells, chromatin state annotation was also performed using fetal kidney and fetal adrenal gland tissues – also made available from the Wang lab². 4C *trans* peaks and signals were also compared to signals for the repressive histone marks H3K9me3, NCBI GEO: GSE66530 (Hattori et al., 2016), and H3K27me3, NCBI GEO: (Lamb et al., 2019), in HEK293 cells and compared to the same marks, NCBI GEO: GSE121984, in iciHHV-6A cells (Saviola et al., 2019).

4C-seq *cis* interactions were determined using the program peakC (Geeven et al., 2018). Telomere content comparison was assessed by TelomereHunter (Feuerbach et al., 2019). Illumina data was simulated using ART (Huang et al., 2012) set to HiSeq2500 data. Reads were produced to cover the entire UCSC hg38 reference genome at 1X coverage. Then, a number of simulated reads equivalent to the number of reads sequenced between 4C data were randomly sampled from the simulated data. TelomereHunter was run on all samples and a non-parametric Wilcoxon test was performed to compare 4C-seq reads to simulated reads.

To identify the chromosomes harboring HHV-6A integration, a scoring system was developed and coded using the R programming language. Accordingly, trimmed reads were aligned to the hg38 genome using bowtie2, or BWA-MEM (Li, 2013) for MinION reads and each read that maps within 500 kb of any autosome terminus was scored. The total scores for reads within these 500 kb regions are tallied and then compared to each other in one of two ways. If replicates are available, a two-way ANOVA is first performed with chromosome end sum scores as the outcome variable and individual chromosome ends as the categorical predictor variable. TukeyHSD pairwise-*post hoc* tests are then performed and the chromosome end with the highest mean $-\log_{10}$ p-value is determined to be the most likely candidate. A significance of 0.05 was used for this approach. If replicates were not available for a specific viewpoint/cell type combination, the chromosome end scores were all compared to each other using Wilcoxon nonparametric tests. The chromosomal end with the largest $-\log_{10}$ p-value adjusted for multiple test comparison was chosen as the most likely candidate and an alpha of 0.05 was used as a cutoff of significance. Furthermore, statistical ranking *via* the aforementioned pairwise post-hoc tests from the algorithm score sums (for each chromosome end) to identify the ends of each chromosome as most significant candidate for HHV-6A integration.

RESULTS

Dynamics of HHV-6A Gene Expression During the Establishment of Latency

We recently reported that the chromosomally integrated HHV-6A genome exists in a compacted transcriptionally silent state

¹https://github.com/FrietzeLabUVM/4c_primer

²https://egg2.wustl.edu/roadmap/web_portal/imputed.html#chr_imp

(Saviola et al., 2019). To study the kinetics of HHV-6A gene silencing upon HHV-6A infection, we performed a 7-day time course in human 293T cells. 293T cells are susceptible to HHV-6A infection and were previously established as a model for studying virus integration (Arbuckle et al., 2010; Gravel et al., 2017; Saviola et al., 2019). Cells were infected with recombinant HHV-6A virus expressing GFP under the control of the major immediate-early (IE) HCMV promoter (Tang et al., 2010a; Wallaschek et al., 2016). Infected GFP-positive cells were isolated by FACS sorting 16 h post infection and subsequently cultured. The presence of the HHV-6A genome was confirmed by qPCR and integration was validated by FISH at day seven post-infection. Infected cells were collected daily until 7 days post-infection (dpi) and processed for gene expression profiling *via* RNA-sequencing (Figure 1A). The expression of all HHV-6A genes progressively decreases over the 7-day period as visualized *via* a clustered heatmap (Figure 1B). The observed gene silencing over this period course can be grouped into three clusters based on expression levels of all viral genes (early, mid and late). The collective expression of genes is reduced over a 3-day period, followed by near complete silencing from 5–7 dpi (Figure 1C). Thus, HHV-6A gene expression upon infection is

silenced within 7 days in 293T cells harboring the integrated virus genome.

Higher-Order Chromatin Interactions of Chromosomally Integrated HHV-6A

The epigenetic mechanisms that regulate gene silencing in chromosomally integrated HHV-6A/B remain poorly characterized. We previously demonstrated that the integrated viral genome forms repressed heterochromatin domains (Saviola et al., 2019). We further hypothesized that higher-order chromatin interactions play a role in virus gene silencing. To investigate higher-order chromatin interactions between chromosomally integrated HHV-6A, we designed 4C-seq assays using distinct viewpoint regions designed against distinct HHV-6A regions (Figure 2A). To obtain the inverse PCR primers required for 4C analysis, we developed a genome agnostic 4C primer design tool (see methods). Two distinct viewpoints (vp1 vs. vp2) were designed based on optimal fragment length and restriction enzyme sequence along the length of the ~160 kb HHV-6A genome. The HHV-6A genomic regions targeted by viewpoint vp1 and vp2 primers are located adjacent to the U39 gene-encoding glycoprotein B (gB) (position: 64 kb) and the U95 gene (position: 148 kb), respectively

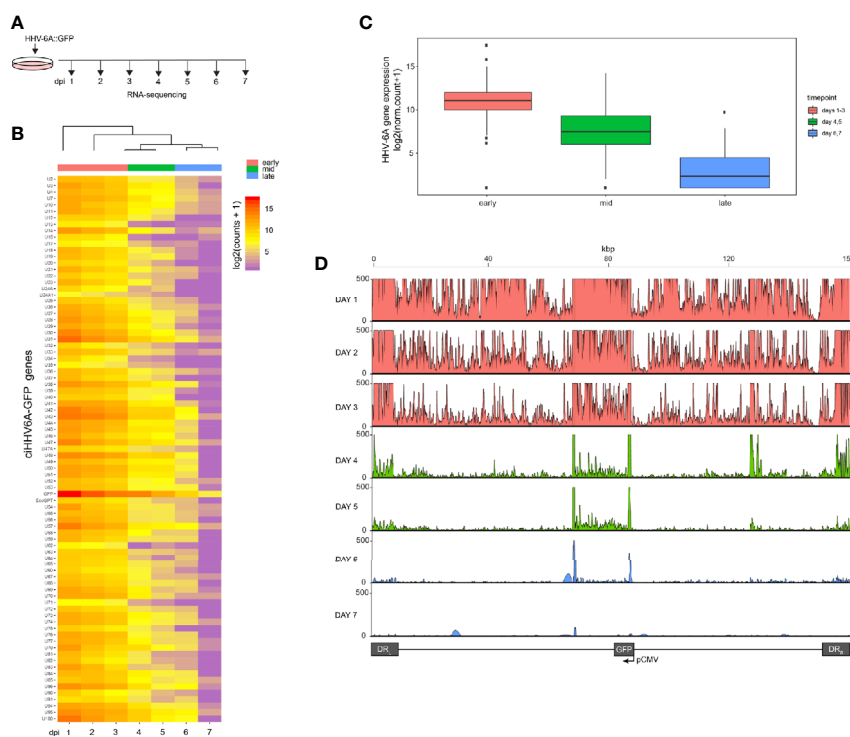


FIGURE 1 | RNA-Seq time course reveals the establishment of transcriptionally silent state upon human herpesvirus 6A (HHV6A) infection. **(A)** 293T cells were infected with ciHHV-6A and green fluorescent protein (GFP) expressing cells were isolated by FACS. RNA-seq analysis was performed on the cells during the integration phase at the indicated days post infection (dpi). **(B)** Cluster analysis of RNA-seq expression of HHV-6A GFP genes, showing expression values of normalized counts with a pseudo-count of one. **(C)** Boxplots of mean viral gene expression shown for each time point. **(D)** RNA-seq signal plotted across the HHV-6A genome for each day. Viral expression is almost completely absent by day 7 indicating the establishment of a transcriptionally silent latent infection. At the later timepoints (e.g., days 4-6), two prominent peaks can be seen flanking the U41 and U42 genes, the origin of lytic DNA replication (oriLyt) and the CMV promoter and GFP regions.

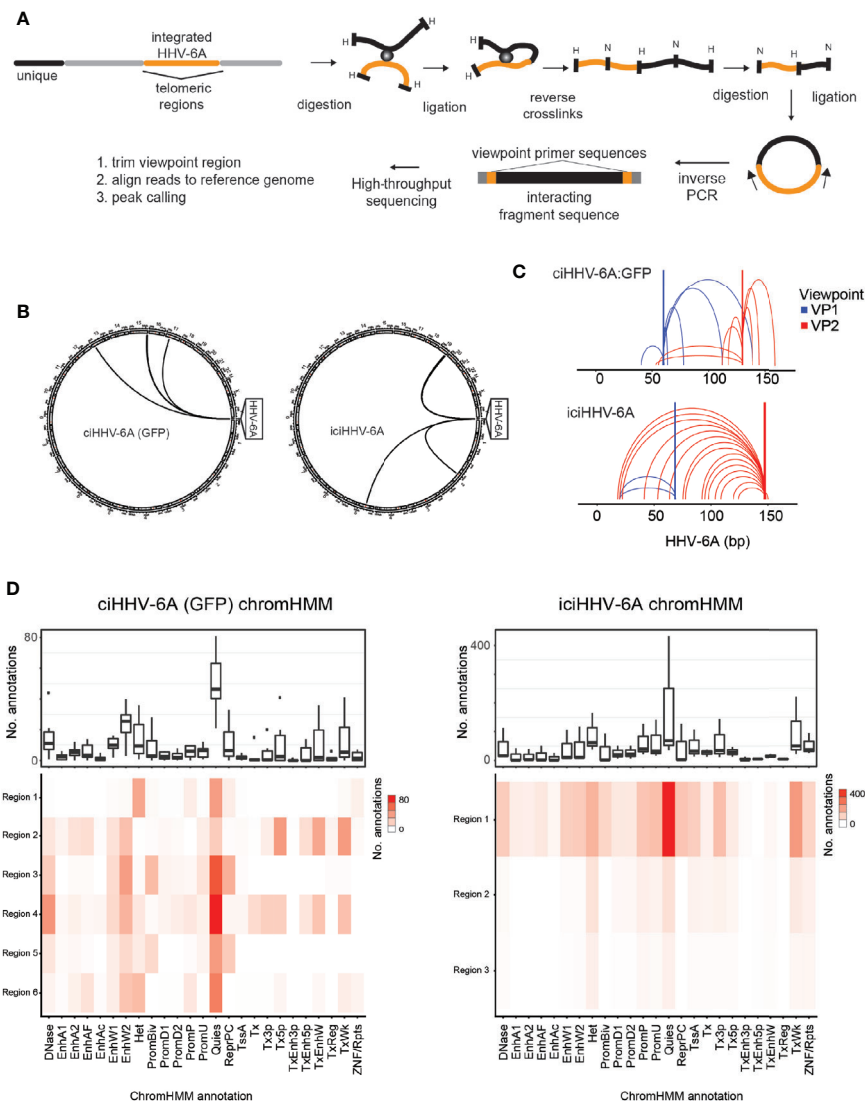


FIGURE 2 | Chromatin interactions of chromosomally integrated human herpesvirus 6A (HHV-6A). Circular Chromosome Conformation Capture (4C-seq) assay was employed using chromosomally integrated HHV-6A cell models to investigate virus-host chromatin interactions. **(A)** 4C-seq assay with ciHHV-6A cells. The integrated HHV-6A genome (yellow) is shown with its flanking telomeric regions (gray) and telomere-proximal unique portions of a human chromosome (black). Indicated are the restriction endonucleases sites used in the 4C-seq assay (HindIII and DpnII) are indicated as ‘H’ and ‘D’, respectively). Inverse PCR was performed with distinct viewpoint primers and analysis interaction regions was performed as shown. **(B)** HHV-6A *trans* interactions for *in vitro* integration model (ciHHV-6A) and the chromosomally inherited *iciHHV-6A* cell model (iciHHV-6A). Significant interactions are visualized using circo plots both viewpoints are shown for ciHHV-6A cells (left) and for *iciHHV-6A* cells (right). **(C)** HHV-6A *cis* interactions identified for two separate HHV-6A viewpoints. Inverse primers were designed for two viewpoint regions chosen at ~ 64 kb and 148 kb along the HHV-6A genome. Note that the HHV6A-GFP genome is slightly larger than the canonical HHV-6A genome due to the additional presence of recombinant elements. **(D)** Host interaction regions (*trans*) are annotated using the 127 epigenomes 25-state chromatin state model database for ciHHV-6A (left) and *iciHHV-6A* (right). The host interaction regions contain the indicated chromatin database annotations. Displayed in boxplots (top) and as a heatmap (below).

(Nicholas and Martin, 1994). With this approach, we assessed the HHV-6A chromatin conformation in a previously described 293T cells line harboring the chromosomally integrated HHV-6A (ciHHV-6A) (Saviola et al., 2019) and in *iciHHV-6A*+ patient-derived cells (*iciHHV-6A*). We obtained approximately 2 hundred thousand to 2 million reads per viewpoint replicate across 2 independent 4C-seq assays, of which 85.93% of reads mapped to

the combined reduced human+HHV-6A genome (hg38 + NC_001664.4 or recombinant GFP genome, depending on the sample) (Supplementary Table 2). Significant ‘*trans*’ interacting regions between the HHV-6A genome and the human genome were identified for both *in vitro*-derived (ciHHV-6A) and patient-derived *iciHHV-6A* cells (Figure 2B). For ciHHV-6A, we found a total of 6 high-confidence interactions that intersected between

viewpoints, and a total of 3 interactions were between viewpoints for *iciHHV-6A* cells. In addition, a number of *cis*-interacting regions were identified for both viewpoints (**Figure 2C**; **Supplementary Table 2**). In line with other 4C-seq results (Krijger et al., 2020) there were relatively high densities of interactions that occur near and between viewpoints. Overall, there were a number of *cis* interactions identified in each cell types suggesting that virus genome may reside in a highly folded compartment similar to a topologically associated domain.

To assess the chromatin features of the identified interaction regions, we annotated the identified interacting human regions using available chromatin state segmentations compiled from multiple human tissue types (Ernst and Kellis, 2015). This annotation dataset is derived from 12 epigenetic features and from 127 original reference epigenomes. We found that the human chromatin that interacted with HHV-6A is largely composed of quiescent and heterochromatin states, whereas enhancer and transcriptional activation annotations occur to a lesser degree (**Figure 2D**). Similarly, chromatin state annotations derived from single representative cell types exhibit similar chromatin state patterns at the HHV-6A interacting genomic regions (**Supplementary Figure 3**). We further inspected the enrichment of the repressive histone modifications H3K9me3 and H3K27me3 at both *ciHHV-6A* and *iciHHV-6A* 4C-seq peaks using ChIP-seq data from each respective cell-line (**Supplementary Figure 2**). In both *ciHHV-6A* and *iciHHV-6A*, the majority of the significant 4C-seq peaks overlap directly with these repressive histone marks. These results indicate that the integrated HHV-6A genome forms higher order chromatin structures within the viral genome as well as with regions of the human genome that are enriched with repressive chromatin. Overall, these results provide insight into HHV-6 chromatin structures within the host cell nucleus.

Proximity Chromatin Ligation Reveals Integration Sites

HHV-6A/B integration sites were first detected by FISH (Nacheva et al., 2008), and the virus-host junction of three different integrations was sequenced using a PCR-based approach. Although PCR amplification and Sanger sequencing provided sequence information, this approach requires previous knowledge of the chromosomal location of the virus genome (Huang et al., 2014; Tweedy et al., 2016) and is prone to amplification biases and sequencing errors due to the repetitive nature of the region (Arbuckle et al., 2010; Arbuckle et al., 2013; Huang et al., 2014; Ohye et al., 2014; Gulve et al., 2017). Genome-wide mapping of short read sequence data to human subtelomeres is challenging due the repetitive nature of subtelomeric regions. It is well-documented that higher order chromatin interactions largely occur in *cis* in large megabase pair (Mbp)-sized topologically associated domains (TADs) (Szabo et al., 2019). Because HHV-6A integrates into a host chromosome, the extra-chromosomal *trans* 4C-seq interactions that are identified should behave like endogenous intrachromosomal *cis* interactions, i.e., 4C-seq should show distinct interacting regions between the host and HHV-6A genomes. With this in mind, we hypothesized that 4C-seq data could identify HHV-6A integration

sites, and visualization of 4C alignments to the human genome reveals particular telomeric ends that are enriched with 4C signal. **Figure 3A** shows enrichment at the distal end of chromosome 15q relative to the rest of the genome in *ciHHV-6A* cells. In addition, clustering of reads at the distal ends of chromosome 15q and chromosome 19q in *ciHHV-6A* and *iciHHV-6A*, respectively, can be seen in **Figure 3B**. Further, we identified significantly more telomeric sequences in our 4C-seq data relative to simulated reads of equal size and GC content (**Supplementary Figure 1**). To enable systematic analysis of HHV-6A integration sites using 4C-seq, we established a scoring procedure to assess the chance of integration across all chromosomes based on clustered read mapping (see methods). The ends of chromosome 15q and chromosome 19q rank as the highest scoring chromosome ends for *ciHHV-6A* and *iciHHV-6A*, respectively, in terms of read mapping density and integration probability (**Figures 3C, D**). This indicates that these two chromosome ends were the most likely candidates for harboring integrated HHV-6A in the *ciHHV-6A* and *iciHHV-6A* cell models, respectively. We confirmed that these chromosomal regions harbor the integrated virus by FISH using probes specific for the virus genome and the respective human chromosomes (**Figure 3E**). The integrated HHV-6A genome indeed was present in the respective chromosomal loci, highlighting that the 4C-seq analysis is an unbiased method to identify HHV-6 chromosomal integration sites.

4C-Seq Integration Site Analysis Using Nanopore Sequencing

Next generation sequencing technologies have made it possible to generate large quantities of sequence data required for a variety of genomic assays including 4C-seq. However, 4C-seq library sizes derived from inverse PCR are relatively large and often difficult to cluster on Illumina flowcells. Long read sequencing platforms, including the Oxford Nanopore Technologies (ONT) minION, have become common an option for long read sequencing and have been applied for structural genome mapping approaches (De Coster et al., 2019). We therefore evaluated minION sequencing on the 4C-seq libraries as a method to identify integration sites from HHV-6 samples. We sequenced the same 4C-seq libraries with minION flowcells and obtained 150,646,076 bp of demuxable reads with a mean read length of 604.0 bp for the *ciHHV-6A* and *iciHHV-6A* sequencing run. For the cell counts titration sequencing run we obtained 419,121,238 reads with a mean read length of 730.4 bp. Alignment of reads that pass quality filters to the human genome reveals a similar grouping of telomere proximal reads at 15q and 19q for *ciHHV-6A* and *iciHHV-6A* samples, respectively (**Figure 4A**). Current 4C-seq peak calling software are not well-suited for long-read with low quality scores. We thus adapted our own method based off clustering of adjacent 4C-seq reads (methods). Using this approach, we found an overall comparable number of *trans* peaks called between replicates (**Figure 4B**). Thus, minION sequencing data represents a sequencing method to generate 4C-seq read data.

We applied the minION 4C-seq reads to investigate the required input material required for identifying HHV-6A

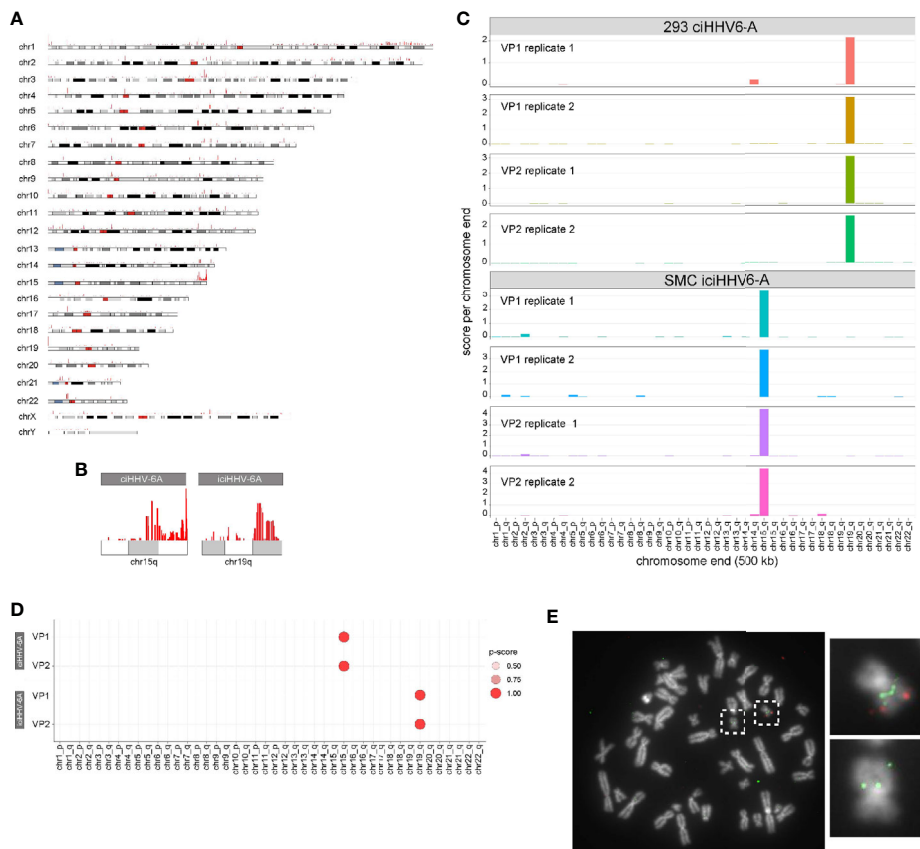


FIGURE 3 | (A) Identification of chromosome ends harboring integrated human herpesvirus 6A (HHV-6A). **(A)** Genome-wide karyotype plot showing 4C-seq read mapping density for iclHHV-6A, viewpoint 6a10 replicate 1, across each human chromosome. **(B)** Karyotype plots showing 4C-seq read mapping density for chromosomes 15 and 19 for each cell-line. **(C)** 4C-seq signal clustering method to score all the reads within 500 kb from each chromosomal end. The sum of these scores is represented in the bar plots for each replicate for each viewpoint in both ciHHV-6A and iclHHV-6A cell types. **(D)** Statistical measure of the likelihood that a specific chromosome harbors integrated HHV-6A. For datasets that contain replicates, an ANOVA was performed using the score from our algorithm across all chromosome ends followed by pairwise Tukey's HSD post-hoc tests. The "p-score" is the mean $-\log_{10}(\text{p-value})$ value for all pairwise comparisons divided by (normalized by) the maximum mean $-\log_{10}(\text{p-value})$. **(E)** Integration sites were confirmed *via* fluorescent *in situ* hybridization using probes complimentary to chr19q (green) and the HHV-6A genome (red) in iclHHV-6A (SMC) cells.

integration sites by titrating the number of cells for 4C library construction. We compared a range of input cell quantities for 4C library construction (10^4 to 10^7 cells) using the iclHHV-6A cells combined with minION sequencing (Figure 4C). These results confirm that 19q harbors the integrated HHV-6A genome for iclHHV-6A SMC cells, and this can be reliably detected in as few as 100,000 cells for the 4C-seq assay. However, using just 10,000 cells 19q also scores highest, however there appears to be some ambiguity as 14q also scores highly. Finally, we also applied minION 4C-seq to detect HHV-6A integration sites using frozen patient-derived B cells from iclHHV-6A+ individuals for this assay. Like the iclHHV-6A SMC samples, 19q also scores as the likely integration site for HHV-6 integration in additional patient-derived lymphocyte samples (Figure 4D). FISH validation using chromosome 19q probes of the patient samples confirms HHV-6A integration into chromosome 19q (Aswad et al., 2020; Wight et al., 2020). Thus, 4C-seq libraries can be generated as little as 10,000 cells and using cryopreserved

lymphocyte cell samples. In summary, this workflow allowed us to efficiently identify the chromosomal ends that most likely harbor the integrated virus genome.

DISCUSSION

HHV-6A and HHV-6B, potentially along with HHV-7 (Prusty et al., 2017) are the only known pathogens to integrate into human telomeres (Gulve et al., 2017; Prusty et al., 2017). While the chromatin mechanisms that govern HHV-6A latency and reactivation remain overall poorly characterized, it has been suggested that integration of the HHV-6A/B genome facilitates the maintenance of the virus genome latency (Arbuckle et al., 2010; Huang et al., 2014). The establishment quiescent/latent states has been investigated in a number of cell lines including 293T cells (Gravel et al., 2017). In these *in vitro*-generated cells and patient-derived iclHHV-6A cells, the virus genome

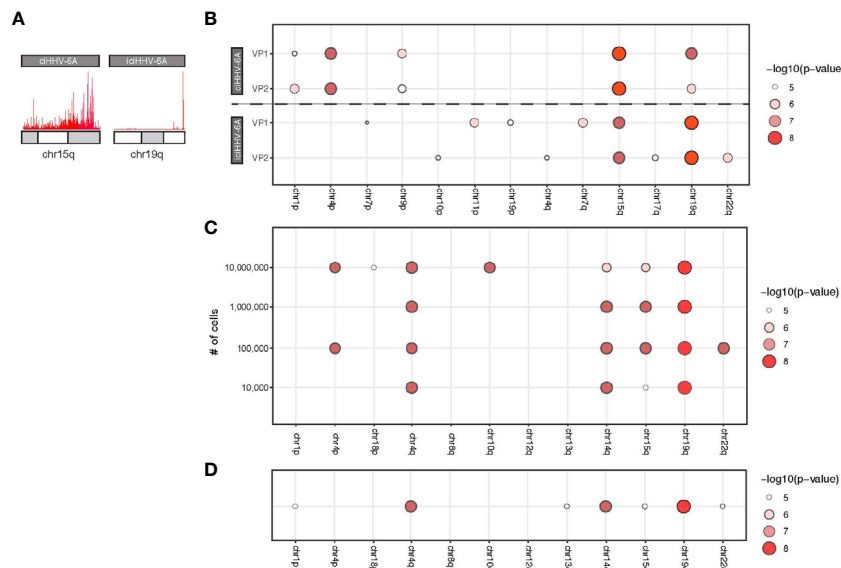


FIGURE 4 | MinION-based 4C-seq human herpesvirus 6A (HHV-6A) integration analysis. **(A)** Karyotype plots showing 4C-seq MinION read mapping density at chr15q and chr19q for ciHHV-6A and iciHHV-6A cell models. **(B)** Chromosomal end integration site analysis using 4C-seq with MinION reads. As in **Figure 3D**, a statistical measure of the likelihood that a specific chromosome harbors integrated HHV-6A. Plotted are only the top scoring chromosome ends. **(C)** A titration series of cell dilutions was used for 4C-seq MinION analysis and was scored with statistical testing to identify potential HHV-6A integration chromosome ends. **(D)** MinION 4C-seq analysis results of a frozen iciHHV-6B+ human lymphocyte sample.

transcriptionally silent (Saviola et al., 2019). These global RNA-sequencing data are in contrast to Kondo et al., who detected a few latency-associated transcripts by qRT-PCR (Kondo et al., 2002).

To investigate the establishment of a latent-like expression profile, we infected cells and performed RNA-sequencing during the first days of integration. Indeed, in early time points (days 1–3, **Figure 1**), we detect high expression levels of all HHV-6A RNA transcripts. However, over the period of 5 days post infection the viral transcripts were significantly reduced and became virtually undetectable by 7 days post infection. It remains unclear whether the RNA-expression is derived from and integrated or extrachromosomal virus genome. However, after 7 days post infection only integrated virus genome were detected when gene expression is absent.

Our prior analyses of the heterochromatin enrichment profiles of integrated HHV-6A genomes revealed that the integrated viral genome was resistant to MNase digestion, suggesting that the viral genome forms a compact chromatin domain (Saviola et al., 2019). Consistent with these findings, they found that the viral genome was enriched with the repressive posttranslational histone modifications H3K27me3 and H3K9me3. These results suggested that repressive chromatin structures likely play a role in viral gene silencing.

To further explore the chromatin-mediated mechanisms associated with silent chromatin, we applied the 4C-seq to determine the genome-wide chromatin contacts of HHV-6A based on nuclear proximity ligation. 4C-seq is designed to examine the genome for sequences contacting a selected genomic site of interest. Accordingly, we designed 4C-seq assays with viral

viewpoint primers to identify the HHV-6 chromatin interactome. We generated high-resolution contact profiles for distinct viral viewpoints and detail the chromosomally integrated HHV-6 genomic organization. We find that clonally expanded *in vitro* HHV-6A (ciHHV-6A) cells display unique virus-host interactions relative to patient-derived iciHHV-6A cells. Overall, there are not many significantly detected inter-chromosomal or *trans* contacts across the human genome. However, the interactions that we detected were largely enriched with chromatin annotated as being quiescent or heterochromatin. Overall, this analysis indicates that the HHV-6A genome folds in a manner that it is capable of physically interacting with repressive chromatin and may explain some of the mechanisms used to silence virus gene expression in chromosomally integrated HHV-6A cells. To our knowledge, these results were the first reported chromatin interaction for a betaherpesvirus, and it is possible that these interactions between the virus and repressive chromatin elements play distinctive functions in HHV-6A latency.

We also found that the integrated HHV-6A genome forms several shared intra-chromosomal or *cis* contacts between the two ciHHV-6A cell types and that the virus genome exists in a highly folded state. It is possible that the HHV-6 genome forms a topologically associated domain within the virus and neighboring human sub-telomeric regions. Topologically associated domains or TADs are large regions of local intrachromosomal interactions. In support of viral TAD formation, we find a strong enrichment of telomeric sequence content in HHV-6A 4C-seq data. Interestingly, these domains appear to extend past the non-unique subtelomeric regions into the unique portion of the human genome and aligned reads cluster

to the ends of 15q and 19q stand out in this data. Importantly, these regions score as the strongest interaction sites and confirm as being the integration chromosomes in independent FISH assays using probes to both virus and the identified chromosomal regions. Further experimental and computational methods including HiC and TAD calling algorithms are required to comprehensively determine HHV-6A TAD structures.

The identified *trans* interacting human sites as well as the strong signals along the distal ends of distinctive human chromosomes indicated that the 4C assay may prove useful to identify integration sites. Current 4C-seq peak calling methods are designed to identify interacting peaks based on monotonic shape as well as proximity to viewpoint regions. In addition, most 4C-seq peak calling methods are not designed to detect *trans* interactions between two unrelated genomes, although recent methodological improvements have been reported in the case of Epstein Barr Virus (Kim et al., 2020) and we were able to successfully apply such a method to our Illumina sequencing data.

To further establish 4C assays for calling integration sites, we developed a simple scoring method that scores and statistically evaluates candidate chromosomal integration sites. This method bins read alignments that occur at the distal 500 kb ends of each chromosome and performs statistical analysis to provide a useful score to evaluate chromosomal integration. For the ciHHV-6A and iciHHV-6A samples, this method accurately predicts the integration site. We further used this method to evaluate reads derived from a minION flow cell from a titration experiment with single replicates for each titration level. In each case, this method accurately identified validated integration regions as top scoring candidate regions and helped to demonstrate that as few as 10,000 – 100,000 cells can be used as input for 4C-seq assays.

Chromatin conformation capture methods have been previously applied to the study of physical chromatin interactions between host and viral genomes. For herpesvirus latency, HiC was used to demonstrate that the latent EBV will disassociate from repressive heterochromatin compartments and form new associations with transcriptionally permissive euchromatin upon reactivation (Moquin et al., 2018). 4C-seq was recently used with Burkitt's lymphoma cells, showing that the latent EBV episomes make contact with transcriptionally repressive H3K9me3 sites as well as attachment sites associated with transcriptionally silent genes (Kim et al., 2020). Finally, 4C-seq has been applied to study the chromatin structures of the latent HIV proviral genome (Dieudonne et al., 2009).

Chromatin conformation capture libraries generally require a large amount of starting material, i.e. as many as 10 million cells (Nilsen, 2014). To enable the 4C assay toward a routine and throughput method to identify integration sites in clinically relevant samples, we performed the assay using a titration of cell numbers ranging from 10^4 to 10^7 cells. Indeed, our titration results demonstrate that we can reliably detect sites as low as 10,000 cells (Figure 4). However, with this low number of cells we identified a few potential spurious hits including chromosome 14. Another adaptation to the assay, is the use of the Nanopore sequencing platform. This newer sequencing platform has several advantages over Illumina, including cost

and speed of data generation. In particular relevance to this study, 4C libraries generated *via* inverse PCR can be greater than 1 kb in size. The resulting Illumina libraries can be difficult to cluster, resulting in a poor output from the sequencing run at a higher cost. We therefore used Nanopore to generate data for integration analysis and find that using our scoring method that this is adequate for integration analysis. Due to the reduced read quality and relatively lower output, it remains to be determined if Nanopore-derived 4C-seq analysis compares to Illumina analysis for interaction peak analysis.

Finally, it is important to note that the resolution of 4C-seq can be as high as 1-2 Mbp for 4C-seq interactions (Krijger et al., 2020) and the annotation of the corresponding physical interactions can be challenging. Higher resolution 3C methods, including capture HiC can be applied in future studies to better study the nature of these identified physical interactions. These issues of scale are particularly important in the case of HHV-6A integration because the interactions between the virus and the host genomes are akin to *cis-trans* interactions. The HHV-6A genome is simultaneously acting like an independent chromosome (*trans*) that integrates into a particular human chromosome (*cis*) and we need to consider these interactions as a special case of 4C interactions, i.e., something like a “*cis-trans*” interaction. Because of this simultaneous “*cis*” nature, we chose to focus on the *trans* interactions in the 1–2 Mb range that is typical of the size of *cis* interactions reported in literature. Future studies aimed at more precisely defining the cellular factors and human contact points made by integrated HHV-6 will facilitate a more mechanistic understanding of viral gene silencing.

In summary, we have utilized a 4C-seq framework to identify both *trans* (virus-host) and *cis* (virus-virus) interactions that are formed within the human host cell nucleus. We further utilized this assay toward the unbiased identification of HHV-6A integration sites in human chromosomes. This optimally complements our recently developed optical mapping approach for the integrated HHV-6 genome (Wight et al., 2020). While research into the mechanisms and factors governing HHV-6A integration is an active area of ongoing research (Aimola et al., 2020), implementation of the methodology presented in this study provides important information on the silencing of the integrated virus genome and lays the foundation for high throughput detection of HHV-6A/B integration sites in clinical samples.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: GSE162821.

ETHICS STATEMENT

Specimens were obtained from the Fred Hutch Research Cell Bank, which prospectively collects and cryopreserves peripheral blood mononuclear cells (PBMCs) from donors and recipients.

The University of Washington Institutional Review Board approved use of the iciHHV-6 specimens from the Fred Hutchinson Cancer Research Center and use of anonymized excess HHV-6-positive samples submitted for testing at the University of Washington Virology lab.

AUTHOR CONTRIBUTIONS

BK and SF designed and supervised the study and acquired funding. CZ and GA performed cell culture, qPCR, FISH, and flow cytometry experiments. MM, EH, PR, AR, and DG performed 4C-seq and RNA-seq experiments. MM and SF performed the analysis of the sequencing data. LF generated and provided cells. MM, AD, and SF wrote the manuscript. MM, BK, SF, LF, and GA revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the NIH R21AI121528, U54GM115516, and European Research Council grant number Stg 677673.

ACKNOWLEDGMENTS

We are grateful to Josh Hill for providing patient-derived cells. We are also grateful to Scott Tighe and Pheobe Kehoe and the

REFERENCES

- Aimola, G., Beythien, G., Aswad, A., and Kaufer, B. B. (2020). Current understanding of human herpesvirus 6 (HHV-6) chromosomal integration. *Antiviral Res.* 176:104720. doi: 10.1016/j.antiviral.2020.104720
- Arbuckle, J. H., Medveczky, M. M., Luka, J., Hadley, S. H., Luegmayr, A., Ablashi, D., et al. (2010). The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 107 (12), 5563–5568. doi: 10.1073/pnas.0913586107
- Arbuckle, J. H., Pantry, S. N., Medveczky, M. M., Prichett, J., Loomis, K. S., Ablashi, D., et al. (2013). Mapping the telomere integrated genome of human herpesvirus 6A and 6B. *Virology* 442 (1), 3–11. doi: 10.1016/j.virol.2013.03.030
- Aswad, A., Aimola, G., Wight, D., Roychoudhury, P., Zimmermann, C., Hill, J., et al. (2020). Evolutionary history of endogenous Human Herpesvirus 6 reflects human migration out of Africa. *Mol. Biol. Evol.* 38 (1), 96–107. doi: 10.1093/molbev/msaa190
- Boyd, J. (2020). *seqsetvis: Set Based Visualizations for Next-Gen Sequencing Data. R package version 1.8.0. ed.* Available at: <https://bioconductor.org/packages/release/bioc/html/seqsetvis.html>.
- De Bolle, L., Naesens, L., and De Clercq, E. (2005). Update on human herpesvirus 6 biology, clinical features, and therapy. *Clin. Microbiol. Rev.* 18 (1), 217–245. doi: 10.1128/CMR.18.1.217-245.2005
- De Coster, W., De Rijk, P., De Roeck, A., De Pooter, T., D'Hert, S., Strazisar, M., et al. (2019). Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res.* 29 (7), 1178–1187. doi: 10.1101/gr.244939.118
- Dieudonne, M., Maiuri, P., Biancotto, C., Knezevich, A., Kula, A., Lusic, M., et al. (2009). Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery. *EMBO J.* 28 (15), 2231–2243. doi: 10.1038/emboj.2009.141
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635

Vermont Integrative Genomics Resource for service. We are also grateful for Jonathan Gordon, Joseph Boyd, Gary Stein, Janet Stein, and Jane Lian for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2021.612656/full#supplementary-material>

Supplementary Figure 1 | HHV-6A 4C-seq reads are enriched with telomeric repeats. Telomere content comparison was assessed by TelomereHunter (Feuerbach et al., 2019). Illumina data was simulated using ART (Huang et al., 2012). Reads were produced to cover the entire UCSC hg38 reference genome at 1X coverage. Then, several simulated reads equivalent to the number of reads sequenced between ciHHV-6A 4C-seq data were randomly sampled from the simulated data. TelomereHunter was run on all samples and a non-parametric Wilcoxon test was performed to compare 4C-seq reads to simulated reads (p-value = 0.002).

Supplementary Figure 2 | HHV-6A *trans* interaction regions are enriched with repressive histone modifications H3K9me3 and H3K27me3. **(A)** Genome tracks across chromosome 15 for ciHHV-6A showing significant 4C *trans* peak regions (top track), 4C signal coverage, HEK293T H3K9me3 ChIP signal and HEK293T H3K27me3 ChIP signal (bottom track). **(B)** The same as **A**, but across chromosome 19 for iciHHV-6A samples.

Supplementary Figure 3 | Similarities in chromatin state annotation in tissues related to HEK293 cells. **(A)** The significant *trans* regions returned by the 4C window method were then annotated using the 127 epigenomes chromHMM 25-state model as in **Figure 2D**. **(B, C)** Annotations for fetal kidney and fetal adrenal gland tissues were then added for comparison with **A**.

- Endo, A., Watanabe, K., Ohye, T., Suzuki, K., Matsubara, T., Shimizu, N., et al. (2014). Molecular and virological evidence of viral activation from chromosomally integrated human herpesvirus 6A in a patient with X-linked severe combined immunodeficiency. *Clin. Infect. Dis.* 59 (4), 545–548. doi: 10.1093/cid/ciu323
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9 (3), 215–216. doi: 10.1038/nmeth.1906
- Ernst, J., and Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* 33 (4), 364–376. doi: 10.1038/nbt.3157
- Feuerbach, L., Sieverling, L., Deeg, K. I., Ginsbach, P., Hutter, B., Buchhalter, I., et al. (2019). TelomereHunter - in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinf.* 20 (1), 272. doi: 10.1186/s12859-019-2851-0
- Flamand, L. (2018). Chromosomal Integration by Human Herpesviruses 6A and 6B. *Adv. Exp. Med. Biol.* 1045, 209–226. doi: 10.1007/978-981-10-7230-7_10
- Geeven, G., Teunissen, H., de Laat, W., and de Wit, E. (2018). peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* 46 (15), e91. doi: 10.1093/nar/gky443
- Gravel, A., Dubuc, I., Wallaschek, N., Gilbert-Girard, S., Collin, V., Hall-Sedlak, R., et al. (2017). Cell Culture Systems To Study Human Herpesvirus 6A/B Chromosomal Integration. *J. Virol.* 91 (14), e00437-17. doi: 10.1128/JVI.00437-17
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics* 30 (19), 2811–2812. doi: 10.1093/bioinformatics/btu393
- Gulve, N., Frank, C., Klepsch, M., and Prusty, B. K. (2017). Chromosomal integration of HHV-6A during non-productive viral infection. *Sci. Rep.* 7 (1), 512. doi: 10.1038/s41598-017-00658-y
- Hattori, T., Lai, D., Dementieva, I. S., Montano, S. P., Kurosawa, K., Zheng, Y., et al. (2016). Antigen clasp by two antigen-binding sites of an exceptionally

- specific antibody for histone methylation. *Proc. Natl. Acad. Sci. U. S. A.* 113 (8), 2092–2097. doi: 10.1073/pnas.1522691113
- Huang, W., Li, L., Myers, J. R., and Marth, G. T. (2012). ART: a next-generation sequencing read simulator. *Bioinformatics* 28 (4), 593–594. doi: 10.1093/bioinformatics/btr708
- Huang, Y., Hidalgo-Bravo, A., Zhang, E., Cotton, V. E., Mendez-Bermudez, A., Wig, G., et al. (2014). Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. *Nucleic Acids Res.* 42 (1), 315–327. doi: 10.1093/nar/gkt840
- Kaufer, B. B., Jarosinski, K. W., and Osterrieder, N. (2011). Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation. *J. Exp. Med.* 208 (3), 605–615. doi: 10.1084/jem.20101402
- Kaufer, B. B. (2013). Detection of Integrated Herpesvirus Genomes by Fluorescence In Situ Hybridization (FISH). *Methods Mol. Biol.* 1064, 141–152. doi: 10.1007/978-1-62703-601-6_10
- Kaufer, B. B., and Flamand, L. (2014). Chromosomally integrated HHV-6: impact on virus, cell and organismal biology. *Curr. Opin. Virol.* 9C, 111–118. doi: 10.1016/j.coviro.2014.09.010
- Kim, K. D., Tanizawa, H., De Leo, A., Vladimirova, O., Kossenkov, A., Lu, F., et al. (2020). Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. *Nat. Commun.* 11 (1), 877. doi: 10.1038/s41467-019-14152-8
- Kondo, K., Shimada, K., Sashihara, J., Tanaka-Taya, K., and Yamanishi, K. (2002). Identification of human herpesvirus 6 latency-associated transcripts. *J. Virol.* 76 (8), 4145–4151. doi: 10.1128/jvi.76.8.4145-4151.2002
- Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E., and de Laat, W. (2020). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* 170, 17–32. doi: 10.1016/j.jmeth.2019.07.014
- Lamb, K. N., Bsteh, D., Dishman, S. N., Moussa, H. F., Fan, H., Stuckey, J. I., et al. (2019). Discovery and Characterization of a Cellular Potent Positive Allosteric Modulator of the Polycomb Repressive Complex 1 Chromodomain, CBX7. *Cell Chem. Biol.* 261365-1379 (10), e1322. doi: 10.1016/j.chembiol.2019.07.013
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi: 10.1186/gb-2009-10-3-r25
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi: 10.1038/nmeth.1923
- Lee, J. W., Meuleman, W., and Kundaje, A. *Imputed signal tracks* (Washington University in St. Louis: Wang Lab). Available at: https://egg2.wustl.edu/roadmap/web_portal/imputed.html#chr_imp (Accessed 27 September 2020).
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Prepr ArXiv:1303.3997*.
- Liao, R., Zhang, R., Guan, J., and Zhou, S. (2014). A New Unsupervised Binning Approach for Metagenomic Sequences Based on N-grams and Automatic Feature Weighting. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 11 (1), 42–54. doi: 10.1109/TCBB.2013.137
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12):550. doi: 10.1186/s13059-014-0550-8
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17 (1), 3. doi: 10.14806/ej.17.1.200
- Moquin, S. A., Thomas, S., Whalen, S., Warburton, A., Fernandez, S. G., McBride, A. A., et al. (2018). The Epstein-Barr Virus Episome Maneuvers between Nuclear Chromatin Compartments during Reactivation. *J. Virol.* 92 (3), e01413-17. doi: 10.1128/JVI.01413-17
- Nacheva, E. P., Ward, K. N., Brazma, D., Virgili, A., Howard, J., Leong, H. N., et al. (2008). Human herpesvirus 6 integrates within telomeric regions as evidenced by five different chromosomal sites. *J. Med. Virol.* 80 (11), 1952–1958. doi: 10.1002/jmv.21299
- Nicholas, J., and Martin, M. E. (1994). Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early gene homologs and transactivating functions. *J. Virol.* 68 (2), 597–610. doi: 10.1128/jvi.68.2.597-610.1994
- Nilsen, T. W. (2014). Preparation of cross-linked cellular extracts with formaldehyde. *Cold Spring Harb. Protoc.* 2014 (9), 1001–1003. doi: 10.1101/pdb.prot080879
- Ohye, T., Inagaki, H., Ihira, M., Higashimoto, Y., Kato, K., Oikawa, J., et al. (2014). Dual roles for the telomeric repeats in chromosomally integrated human herpesvirus-6. *Sci. Rep.* 4:4559. doi: 10.1038/srep04559
- Okuno, T., Takahashi, K., Balachandra, K., Shiraki, K., Yamanishi, K., Takahashi, M., et al. (1989). Seroepidemiology of human herpesvirus 6 infection in normal children and adults. *J. Clin. Microbiol.* 27 (4), 651–653. doi: 10.1128/jcm.27.4.651-653.1989
- Osterrieder, N., Wallaschek, N., and Kaufer, B. B. (2014). Herpesvirus Genome Integration into Telomeric Repeats of Host Cell Chromosomes. *Annu. Rev. Virol.* 1 (1), 215–235. doi: 10.1146/annurev-virology-031413-085422
- Prusty, B. K., Gulve, N., Rasa, S., Murovska, M., Hernandez, P. C., and Ablashi, D. V. (2017). Possible chromosomal and germline integration of human herpesvirus 7. *J. Gen. Virol.* 98 (2), 266–274. doi: 10.1099/jgv.0.000692
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi: 10.1093/bioinformatics/btq033
- Saviola, A. J., Zimmermann, C., Mariani, M. P., Signorelli, S. A., Gerrard, D. L., Boyd, J. R., et al. (2019). Chromatin Profiles of Chromosomally Integrated Human Herpesvirus-6A. *Front. Microbiol.* 10:1408. doi: 10.3389/fmicb.2019.01408
- Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci. Adv.* 5 (4), eaaw1668. doi: 10.1126/sciadv.aaw1668
- Tang, H., Kawabata, A., Yoshida, M., Oyaizu, H., Maeki, T., Yamanishi, K., et al. (2010a). Human herpesvirus 6 encoded glycoprotein Q1 gene is essential for virus growth. *Virology* 407 (2), 360–367. doi: 10.1016/j.virol.2010.08.018
- Tang, H., Sadaoka, T., and Mori, Y. (2010b). [Human herpesvirus-6 and human herpesvirus-7 (HHV-6, HHV-7)]. *Uirusu* 60 (2), 221–235. doi: 10.2222/jvs.60.221
- Tweedy, J., Spyrou, M. A., Pearson, M., Lassner, D., Kuhl, U., and Gompels, U. A. (2016). Complete Genome Sequence of Germline Chromosomally Integrated Human Herpesvirus 6A and Analyses Integration Sites Define a New Human Endogenous Virus with Potential to Reactivate as an Emerging Infection. *Viruses* 8 (1), 19. doi: 10.3390/v8010019
- van de Werken, H. J., de Vree, P. J., Splinter, E., Holwerda, S. J., Klous, P., de Wit, E., et al. (2012). 4C technology: protocols and data analysis. *Methods Enzymol.* 513, 89–112. doi: 10.1016/B978-0-12-391938-0.00004-5
- Wallaschek, N., Sanyal, A., Pirzer, F., Gravel, A., Mori, Y., Flamand, L., et al. (2016). The Telomeric Repeats of Human Herpesvirus 6A (HHV-6A) Are Required for Efficient Virus Integration. *PLoS Pathog.* 12 (5), e1005666. doi: 10.1371/journal.ppat.1005666
- Wight, D. J., Aimola, G., Aswad, A., Jill Lai, C. Y., Bahamon, C., Hong, K., et al. (2020). Unbiased optical mapping of telomere-integrated endogenous human herpesvirus 6. *Proc. Natl. Acad. Sci. U. S. A.* 117 (49), 31410–31416. doi: 10.1073/pnas.2011872117
- Zerr, D. M., Meier, A. S., Selke, S. S., Frenkel, L. M., Huang, M. L., Wald, A., et al. (2005). A population-based study of primary human herpesvirus 6 infection. *N. Engl. J. Med.* 352 (8), 768–776. doi: 10.1056/NEJMoa042207
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9 (9), R137. doi: 10.1186/gb-2008-9-9-r137

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mariani, Zimmerman, Rodriguez, Hasenohr, Aimola, Gerrard, Richman, Dest, Flamand, Kaufer and Frietze. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.