DNA virus infections shape transposable elements activity *in vitro* and *in vivo*

Jiang Tan[1], Vedran Franke[2], Eva Neugebauer[1,3,4], Justine Lagisquet[5], Anna Katharina Kuderna[6], Stephanie Walter[5], Armin Ensser[5], Markus Landthaler[2,7], Thomas Stamminger[6], Thomas Gramberg[5], Altuna Akalin[2], Emanuel Wyler[2] and Florian Full[1*]

[1] Institute of Virology, University Medical Center, and Faculty of Medicine, Albert-Ludwig-University Freiburg, Freiburg, Germany

[2] Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Helmholtz Society, Berlin, Germany

[3] Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg

[4] Faculty of Biology, University of Freiburg

[5] Institute of Clinical and Molecular Virology, Universitätsklinikum Erlangen, Friedrich-Alexander Universität Erlangen-Nürnberg, 91054 Erlangen, Germany

[6] Institute of Virology, Ulm University Medical Center, Ulm, Germany

[7] Institute for Biology, Humboldt University of Berlin, Berlin, Germany

*Correspondence to:

Florian Full, PhD

Institute of Virology

University Medical Center Freiburg

Hermann-Herder-Str. 11

79104, Freiburg, Germany

florian.full@uniklinik-freiburg.de

Abstract

Transposable elements (TEs) are implicated in a variety of processes including placental and preimplantation development and a variety of human diseases. TEs are known to be activated in the context of viral infections, but the mechanisms and consequences are not understood.

We show strong activation of TEs upon DNA virus infection, in particular the MLT- and THE1-class of LTR-containing retrotransposons as well as a subset of LINE-1-, Alu-elements and HERVs. Mechanistically, two pathways induce TE upregulation: inhibition of the KAP1/TRIM28 repressive complex by phosphorylation, and expression of the pioneer factor double-homeobox 4 (DUX4), which is known to be crucial for TE-induction during zygotic genome activation in embryonic development. In adults, DUX4 is usually silenced and we showed DUX4 expression upon infection with various DNA viruses. DUX4 binds to TEs upon infection and analysis of genes adjacent to TEs shows pathways that are important for DNA virus infections. Analysis of knockdown, knockout and overexpression data reveal that almost all TEs expressed upon herpesviral infection are regulated by KAP1/TRIM28 and DUX4. Interestingly, analysis of single cell sequencing data from patients with DNA virus-associated cancers showed that in vivo TE expression strongly correlates with virus infection, indicating a possible role in viral oncogenesis.

Key words: Transposable Elements, DNA virus, HSV-1, HCMV, KAP1, TRIM28, DUX4, LINE-1, LTR

Introduction

Transposable elements (TEs) are DNA segments that have the ability to move within the genome (1). TEs are reminiscent of past virus infections that resulted in integration into the host genome but lost the ability to produce virus particles and to propagate. The human genome is composed of approximately 45% TEs (2). TEs can be classified into two categories: class-I transposons (retrotransposons) and class-II transposons (DNA transposons). DNA transposons contain a transposase and have a mobile DNA intermediate, while retrotransposons require an RNA intermediate and a reverse transcriptase (3).

Retrotransposons are of scientific interest due to their use of an RNA intermediate and self-replicating life cycle for dissemination, similar to retroviruses. Retrotransposons are classified by function and structure into LINE (long interspersed nuclear elements), SINE (short interspersed nuclear elements), and LTR (long terminal repeats) (4). Once dismissed as parasitic or 'junk' DNA, they have emerged as integral components with functional significance within host genomes. Proteins originally encoded by retrotransposons have been co-opted for regulating essential processes, including placental cytotrophoblast fusion, preimplantation development, and intracellular RNA transport across neurons (5–7), a process named retrotransposon exaptation. Retrotransposon exaptation significantly expands mechanisms of gene regulation and transcript diversity of host genes by enriching them with numerous cis-regulatory elements. This is particularly true for retrotransposons harboring intact LTR elements with intrinsic promoter and enhancer activities, as well as splicing donor/acceptor sequences (8–11).

Although retrotransposons can be threats to genome integrity, and are therefore generally inactivated through degenerative mutations or epigenetic silencing, a subset of them are robustly induced and tightly regulated in specific developmental,

physiological, and pathological contexts. These contexts include preimplantation development, germ cell development, immune response, aging, and cancer (12–17). Proteins like SETDB1 or SUMO-modified KAP1 (TRIM28), associated with histone methylation and Krüppel-associated box domain (KRAB) zinc finger proteins (ZNFs), have been implicated in the repression of TEs (18). In addition, it is known that multiple TEs are activated during zygotic genome activation (ZGA) in early embryonic development. ZGA is driven by the pioneer factor double homeobox 4 (DUX4), which is exclusively expressed during the 4-cell to 8-cell state in human embryonic development (19). Certain viral infections, such as SARS-CoV-2, Influenza A virus, HIV-1 and DNA viruses have been shown to induce the expression of retrotransposons, including endogenous retroviruses (ERVs). This suggests a potential link between viral infection and retrotransposon expression (20–23), however the molecular mechanisms of TE-activation still remain poorly understood.

We recently discovered that DUX4 is upregulated during herpesviral infection (24), making it an excellent model for exploring the mechanisms of controlled induction of TEs. To address these gaps, our goal was to comprehensively analyze TE-transcription upon herpesvirus infection. Additionally, we examined single cell RNA-sequencing (RNA-seq) data of cancer patients that suffer from tumors caused by DNA viruses, namely Merkel cell polyomavirus, human papillomavirus, and Epstein-Barr virus, confirming *in vivo* the induction of TEs in viral cancers. We demonstrate that HSV-1 infection induces the phosphorylation of KAP1, which is found to bind to the DUX4 promoter, thereby modulating the expression of both DUX4 and TEs. Phosphorylation of KAP1 has been shown to regulate sumoylation and KAP1 repression negatively (25). This implies that viral infection leads to TEs expression through regulation of KAP1 and induction of DUX4. This investigation explores the complex relationship between DNA viruses, germline transcription factor DUX4, and

TEs and provides new insights into the regulatory networks that control TEs activation.

Materials and Methods

Cell culture

HEK 293T and human foreskin fibroblasts (HFF) cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Thermo Fisher Scientific), supplemented with 10% (vol/vol) heat-inactivated fetal bovine serum (FBS, Thermo Fisher Scientific), 2 mM GlutaMAX (Thermo Fisher Scientific), 1 mM HEPES (Thermo Fisher Scientific), and 1% penicillin-streptomycin (vol/vol). The maintenance medium for DUX4 inducible cells consisted of DMEM supplemented with 10% (vol/vol) FBS, 2 mM GlutaMAX (Thermo Fisher Scientific), 10 mM HEPES buffer (Thermo Fisher Scientific), 1% penicillin-streptomycin (vol/vol), and 1 µg/ml Puromycin. All cell lines were cultured under standard conditions, and regular monitoring for mycoplasma contamination was performed monthly using the MycoAlert Kit from Lonza. DUX4 overexpression was initiated by the addition of 1 µg/ml doxycycline to the culture medium.

Generation of shKAP1 HFF cell lines and DUX4 inducible 293T cell lines

For the Generation of shKAP1 HFF cell lines, 5 x 10^6 HEK 293T cells were seeded in 10-cm dishes in Dulbecco's minimal essential medium (DMEM) (Gibco) supplemented with 10 % fetal calf serum (FCS, Sigma-Aldrich), and 1 % penicillin-streptomycin (PS, Sigma-Aldrich). Utilizing the Lipofectamine 2000 reagent (Invitrogen), the cells were co-transfected on the next day with packaging plasmids pLP-VSVg (26) and pCMVdeltaR8.9 (kindly provided by T. Gramberg) together either with the plasmid pAPM-D4-miR30-L1221 (a gift from Jeremy Luban; Addgene #115846) (27) expressing a control shRNA or plasmid pAPM-D4 miR30-TRIM28 ts3 (Addgene #115864) (27) encoding for a shRNA targeting KAP1. 48 hours post transfection, the

supernatant was harvested, filtered through a 0.45-µm filter and used for transduction of HFF cells. This was done by seeding HFF in 6-wells (8 x 10^4/well) and incubating them for 24 h with 50 µl of the regarding lentiviral supernatant and 7.5 µg/mL polybrene (Sigma-Aldrich). To select for successfully transduced HFF, the cells were then provided with medium (DMEM, 10 % FCS, PS) containing 5 µg/ml puromycin. The knockdown of KAP1 was verified via western blotting and immunofluorescence. For the Generation of DUX4 inducible 293T cell lines, the processes are similar to above. Transfections were carried out using GenJet (SignaGen Laboratories) or Lipofectamine 2000 (Thermo Fisher Scientific) following the respective manufacturer's protocols. The DUX4-inducible plasmid pCW57.1-DUX4-CA was obtained from Addgene (#99281).

Western Blotting

Cells were lysed using RIPA HS buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM NaCl, 1% Triton X-100 (vol/vol), 0.1% SDS (vol/vol), 0.1% deoxycholic acid (DOC)), supplemented with Aprotinin and Leupeptin, MG-132, and sodium metavanadate (Sigma-Aldrich). The cell pellet was subjected to centrifugation at 4°C, 14,000 rpm for 30 minutes. Subsequently, the resulting samples were diluted with a Laemmli-SDS sample buffer and heated for 5 minutes at 95°C.

RNA-seq library preparation

HFF/shCtrl and HFF/shKAP1 were seeded in duplicates in 6-wells (3 x 10^5/well) in medium without selection (DMEM + 10 % FSC + PS). The day after, the medium was harvested and stored at 37 °C. The cells were either mock infected (1.5 ml fresh medium) or infected with the HCMV laboratory strain TB40/E wild type (28) (1.5. ml medium/virus mix) to obtain an multiplicity of infection (MOI) of 1. The viral titer was

defined based on immediate early (IE) protein-forming units as described elsewhere (29). 2 hours post infection (hpi) the cells were provided with additionally 1.5 ml of the previously conditioned medium. 48 hpi the cells were harvested with 700 µl TRIzol Reagent (Invitrogen) per well and stored at -80°C. For the library of DUX4 inducible 293T cell lines, DUX4 overexpression was initiated by the addition of 1 µg/ml doxycycline for 12 h. RNA was isolated using the Direct-zol RNA Miniprep Kit (Zymo Research) and further subjected to rRNA depletion using the riboPOOL kit (Biozym). To clean up the RNA in the process of rRNA depletion, AMPure beads XP (Beckman Coulter) were utilized. Afterwards, 5 ng of the resulting RNA was applied for library preparation using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (New Englands Biolabs).

Genes and TEs expression analysis in Bulk RNA-seq

For bulk RNA-seq. analysis we employed the snakePipes mRNA-seq pipeline (30) to process the sequencing data. Adapters and low-quality bases were removed using TrimGalore (https://github.com/FelixKrueger/TrimGalore). The resulting trimmed reads were aligned to the human genome (hg38) using the STAR aligner (31). Subsequently, featureCounts (32) was utilized to count the aligned reads, generating a count matrix for downstream analyses. To visualize the coverage, Bigwig files were generated through deepTools bamCoverage (33), employing size factors calculated using DESeq2 (34). For analysis of TEs expression, the snakePipes noncoding-RNA-seq pipeline (30) was employed. Similar to gene expression analysis, preprocessing steps included the removal of adapters and low-quality bases using TrimGalore. The trimmed reads were aligned to the Repeatmasker repeat annotation, obtained from the UCSC database, utilizing the STAR aligner (31). To quantify TEs expression, TEtranscripts (35) was employed to count aligned TE reads. Subsequently, the expression values

were normalized and the differential expression of TEs were analyzed using DESeq2 (34).

Genes and TEs expression analysis in single cell RNA-seq

For the analysis of single-cell RNA-seq data obtained from patients, we employed the STARsolo pipeline (36) to align sequencing reads to the human reference genome (hg38) and corresponding viral genomes, including EBV (NC_007605), HPV (NC_001526.4), and MCPyV (NC_010277). The generated feature-barcode matrices from all individual cells were aggregated and transformed into a Seurat object using the R package Seurat (37). For the analysis of TEs at the single-cell level, we utilized the scTE package (38). TEs expression matrices from individual cells were aggregated and transformed into a Seurat object using the same Seurat package (37). We then merged the gene expression matrices and TEs expression matrices. This integrated dataset was subjected to log-normalization and linear regression using the NormalizeData and ScaleData functions of the Seurat package. Principal component analysis (PCA) was then performed, and the first 15 principal components (PCs) were utilized for Unifold Manifold Approximation and Projection (UMAP) non-linear dimensionality reduction. This was followed by the calculation of a k-nearest neighbor graph and Louvain clustering. Marker genes per cluster were identified using the FindAllMarkers() function, based on the Wilcoxon Rank Sum test. To visualize gene expression patterns, the DoHeatmap() function was employed to generate heatmaps for the top variable genes, based on scaled expression values. Marker gene expression and SingleR analysis (39) were used for annotating cell-type-specific clusters. Clusters with concurrent annotations were identified and merged to generate the cell-type annotated object. For additional visualization, FeaturePlot(), DotPlot() and Vlnplot() functions were utilized to explore genes and TEs expression. Statistical

significance was determined using two-sided nonparametric Wilcoxon rank-sum tests (*P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001).

ChIP-seq analysis

For ChIP-seq analysis, we processed the data using the snakePipes DNA-mapping and the ChIP–seq pipelines (30). Adapters and low-quality bases were then removed employing TrimGalore (https://github.com/FelixKrueger/TrimGalore). The resulting trimmed reads underwent alignment to the human genome (hg38) using Bowtie2 (40). Reads mapping to blacklisted regions from the Encode Consortium69 were excluded from further analysis. Additionally, Picard MarkDuplicates (v.1.65; https://broadinstitute.github.io/picard/) was utilized to identify and remove duplicated reads. For subsequent peak calling analysis, only properly paired mapped reads and reads with a mapping quality over 3 were retained. Peak calling was executed using MACS2 (41), employing the input as a control for comparison. To assess log2(FC) on peak regions between mock and HSV-1 infection, CSAW (42) was employed. To visualize the ChIP-seq data, Bigwig files were generated using deepTools bamCoverage (33). To identify peaks overlapping with TEs, the obtained peaks were intersected with the Repeatmasker repeat annotation downloaded from the UCSC database.

ATAC-seq analysis

For ATAC-seq analysis, data processing was conducted using the snakePipes DNA-mapping and the ATAC–seq pipelines (30), with the DNA-mapping segment aligning data similar to the CHIP–seq analysis (see above). To ensure data quality, TrimGalore (https://github.com/FelixKrueger/TrimGalore) was employed to remove adapters and low-quality bases. The resulting trimmed reads were aligned to the human genome

(hg38) using Bowtie2 (40). Reads mapping to blacklisted regions from the Encode Consortium69 were excluded from subsequent analysis. Additionally, duplicated reads were identified and filtered out using Picard MarkDuplicates (v.1.65; https://broadinstitute.github.io/picard/). For the ATAC-seq analysis, BAM files were further filtered to include only properly paired reads with appropriate fragment sizes (<150 bases). To identify accessible chromatin regions, peak calling was performed using MACS2 (41). CSAW (42) was also applied to calculate the log2(FC) on peak regions between mock and HSV-1 infection. For data visualization, Bigwig files were generated using deepTools bamCoverage (33). To investigate peaks overlapping with TEs, the obtained peaks were intersected with the Repeatmasker repeat annotation downloaded from the UCSC database.

Data visualization

The representation of overlapping genes and TEs across diverse samples was visually conveyed through Venn diagrams, generated using the R package VennDiagram (43). Expression patterns of DUX4 target genes were effectively depicted via heatmap created with the R-package pheatmap (https://davetang.github.io/muse/pheatmap.html). Binding patterns of DUX4 were depicted via a pieplot created with the R package webr (https://github.com/cardiomoon/webr/tree/master/R). Differential expression of genes and TEs was visually summarized through violin plots, donut plots, and volcano plots, all generated using the R package ggplot2 (44). Metagene profiles and heatmaps for both ATAC–seq and ChIP–seq data were constructed using deepTools' plotProfile and plotHeatmap (33). While representative tracks were produced utilizing pyGenomeTracks (45). To gain insights into the functional implications, KEGG pathways enrichment analysis was performed using ClusterProfiler (46).

Statistical analysis

P-values were computed utilizing an unpaired Student's t-test, with statistical significance defined as $P < 0.05$.

Results

TEs are strongly induced following DNA viruses infection

To analyze expression of TEs in response to herpesvirus infection, we generated RNA-sequencing data derived from HFF cells upon HCMV infection and conducted an in-depth analysis of published RNA-seq datasets derived from HFF cells both pre- and post-exposure to HSV-1 for 2, 4, 6, 8 hours, and HCMV for 48 hours. We observed significant upregulation of numerous TEs upon herpesviral infection (Fig.1A,B). TEs-activation became evident between 2 and 4 hours post-infection (hpi) with HSV-1. Notably, from 6 hpi onwards, the upregulated TEs surged dramatically, with about 1000 TEs significantly upregulated at 8 hpi (Fig.1A,B). TEs of all subclasses were upregulated in the course of infection with the numbers at 8 hpi distributed as follows: 208 (DNA), 157 (LINE), 58 (SINE), and 532 (LTR) (Fig.1C; Fig.S1). In the context of HCMV infection, 554 upregulated TEs were induced at 48 hpi (Fig.1D,B), and remarkably, 96.2% of these upregulated TEs (533) exhibited overlap with the upregulated TEs identified in HSV-1 infection at 8hpi (Fig.1E). This observed convergence underscores shared regulatory mechanisms in the TEs response to herpesviral infection.

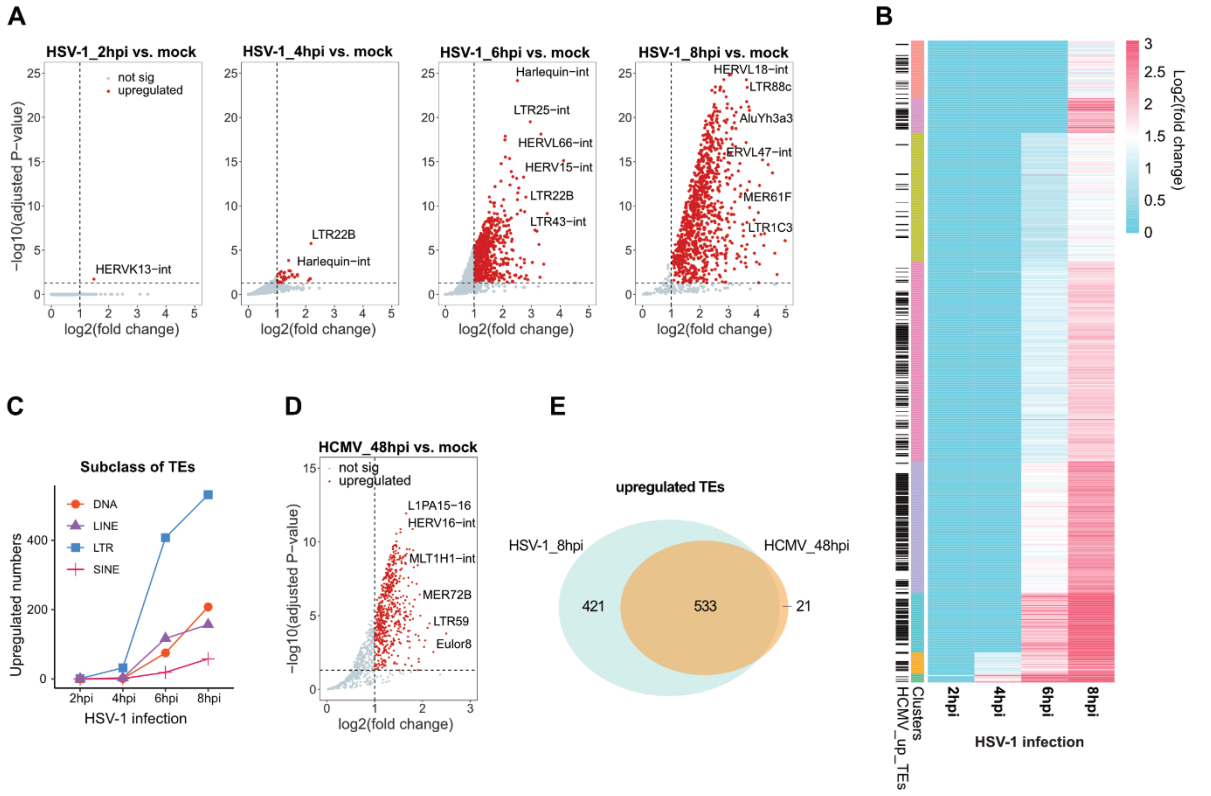

Figure 1: Comprehensive Analysis of TEs in Viral Infections. (**A**) Volcano plots showing significantly upregulated TEs upon HSV-1 infection of HFF cells at 2, 4, 6, and 8 hpi. Red dots indicate TEs meeting the criteria of log2(FC) > 1 and P < 0.05. (**B**) Heatmap depicting the log2(FC) of upregulated TEs at 2, 4, 6, and 8 hours post HSV-1 infection (hpi) compared to mock that cluster in 8 clusters. The significantly upregulated TEs upon HCMV infection at 48 hpi are shown with black in HCMV_up_TEs. Values exceeding log2(FC) > 3 are capped at 3. (**C**) Quantification of the total number of upregulated TEs families per TEs subclass following HSV-1 infection at 2, 4, 6, and 8 hpi. (**D**) Volcano plot highlighting significantly upregulated TEs upon HCMV infection of HFF cells at 48 hpi. Red dots indicate TEs meeting the criteria of log2(FC) > 1 and P < 0.05. (**E**) Venn diagram showing the overlap of upregulated TEs induced by HSV-1 in HFF cells at 8 hpi and HCMV at 48 hpi, providing insights into shared regulatory elements across different viral infections.

TEs contribute to dynamic chromatin regions in response to HSV-1 infection

Next, we analyzed chromatin accessibility and histone modification patterns during HSV-1 infection to investigate the epigenetic regulation of TEs in response to herpesvirus infection. Employing transposase-accessible chromatin sequencing (ATAC-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) targeting the transcriptional repression marks H3K9me3 and H3K27me3, we found dynamic changes in the epigenetic landscape. In uninfected cells, approximately 100,000 peaks

were identified through ATAC-seq. Following HSV-1 infection, the number of peaks exhibited a declining trend over time, reaching around 60,000 peaks at 8 hpi (Fig.S2A). Notably, 18 % - 23 % of these peaks were associated with TEs, a proportion that decreased post-infection (Fig.S2A). To further investigate TEs with altered accessibility in response to HSV-1 infection, we compared ATAC-seq peaks in infected and uninfected samples. The analysis revealed approximately 3,000, 12,000, and 9,000 peaks, marking sites with enhanced accessibility at 4, 6, and 8 hpi, respectively (Fig.2A,B). At 8 hpi, these upregulated ATAC peaks in repeats were distributed across 437 TEs families, with 98.4% of them exhibiting an upregulated TEs expression (Fig.2C,D).

Considering that H3K9me3 and H3K27me3 are associated with transcriptional repression, we compared ChIP-seq peaks for these histone markers during the later stages of HSV-1 infection versus 1 hpi (47). Notably, a substantial number of H3K9me3 and H3K27me3 markers exhibited downregulation at 8hpi (Fig.S2B). For instance, TEs families like the retrotransposons MLT1A and MLT1A0 showcased these distinct epigenetic patterns (Fig.2E,F; Fig.S2C). Taken together, these results show that HSV-1 infection regulates the epigenetic landscape of the human genome, particularly within TEs. The observed changes in chromatin accessibility and histone modifications at TEs suggest a potential regulatory role for these elements in the host response to viral infection.

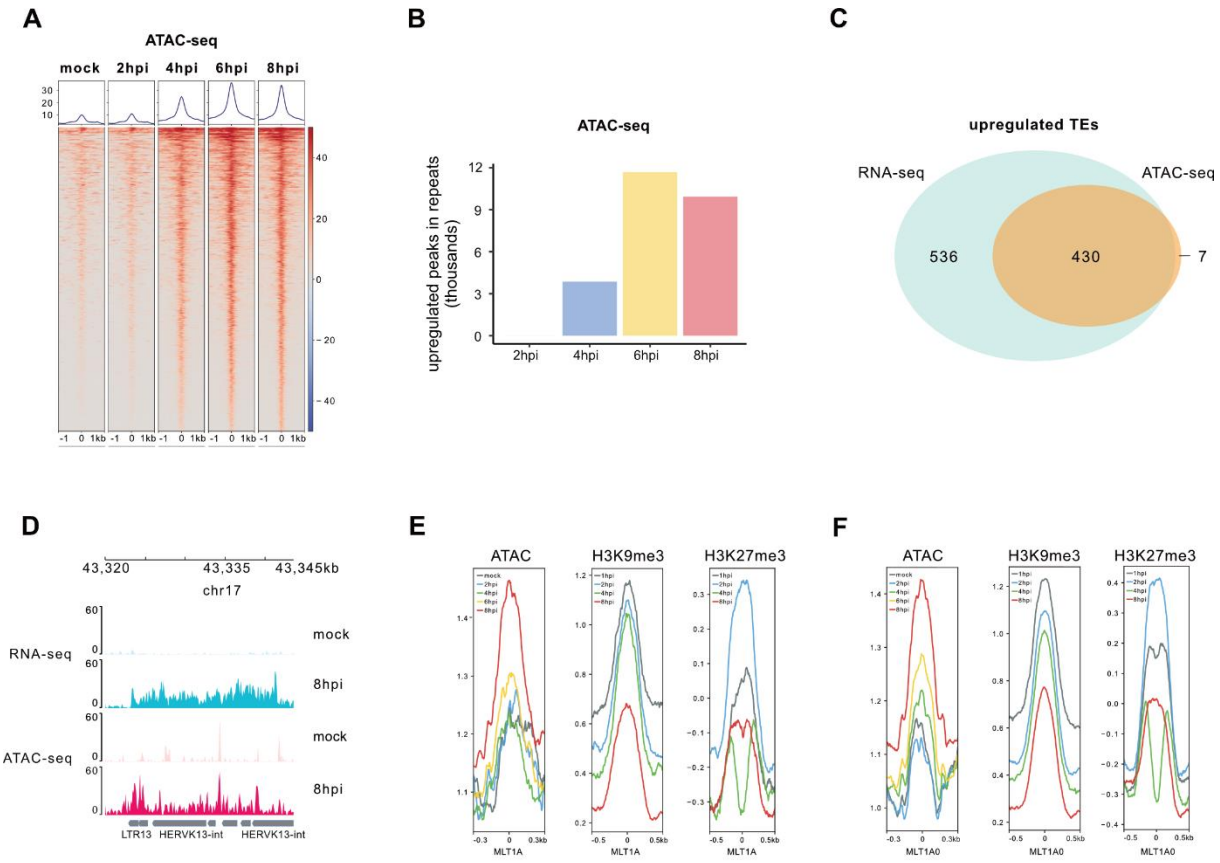

Figure 2: Comprehensive Analysis of Chromatin Accessibility Dynamics during HSV-1 Infection. (**A**) Heatmap displaying enrichment levels of TEs with enhanced accessibility in ATAC-seq. upon HSV-1 infection of HFF cells at mock, 2, 4, 6, and 8 hours post-infection (hpi). (**B**) Number of upregulated ATAC-seq. peaks in repeats detected in HSV-1-infected samples (HSV-1 vs. mock). (**C**) Venn diagram illustrating the overlap of upregulated ATAC-seq. peaks in 437 TEs families and upregulated TEs expression induced by HSV-1 at 8hpi. (**D**) Example genomic view illustrating TEs with upregulated expression and ATAC-seq. peaks at 8hpi. (**E**) Peak count frequency of ATAC-seq, ChIP-seq of transcriptional repression marks (H3K9me3 and H3K27me3) peaks overlapping with MLT1A. (**F**) Peak count frequency of ATAC-seq, ChIP-seq of transcriptional repression marks (H3K9me3 and H3K27me3) peaks overlapping with MLT1A0.

The transcription factor DUX4 binds to TEs

In a previous publication, we identified the germline transcription factor DUX4 as being exclusively induced by DNA viruses and crucial for herpesvirus infection (24). DUX4 is known to play a role in the activation of TEs during ZGA. We therefore sought to determine whether DUX4 binding to TEs regulates their expression and that of nearby genes both in the context of HSV-1 infection and ectopic overexpression. Upon reanalyzing DUX4 ChIP-seq data from both HSV-1-infected cells and DUX4-induced

cells, we observed that 69% and 62% of the binding peaks, respectively, were located within TEs (Fig.3A,B,C). A substantial portion of these DUX4-binding TEs belonged to the MLT and THE families, with 1709 DUX4 binding sites at TEs overlapping between HSV-1-infected cells and DUX4-induced cells (Fig.3D). Further exploration involved the analysis of ATAC-seq and ChIP-seq data for the transcriptional repression marks H3K9me3 and H3K27me3 following HSV-1 infection. Remarkably, within the DUX4-binding TEs, we observed heightened accessibility, while H3K9me3 and H3K27me3 peaks were diminished, particularly at 8 hpi (Fig.3E; Fig.S2C). These findings underscore a dynamic interplay between DUX4, TEs, and histone modifications in response to HSV-1 infection. TEs are known to act as cis-regulatory elements regulating gene expression. Therefore we examined the genes proximal to DUX4-binding sites within TEs and noted a significant enrichment of genes in virus-related pathways, including those associated with KSHV, HPV, HCMV, and HSV-1 infection (Fig.3F, Fig.S2D), indicating a role of DUX4 binding and TE expression in virus infection.

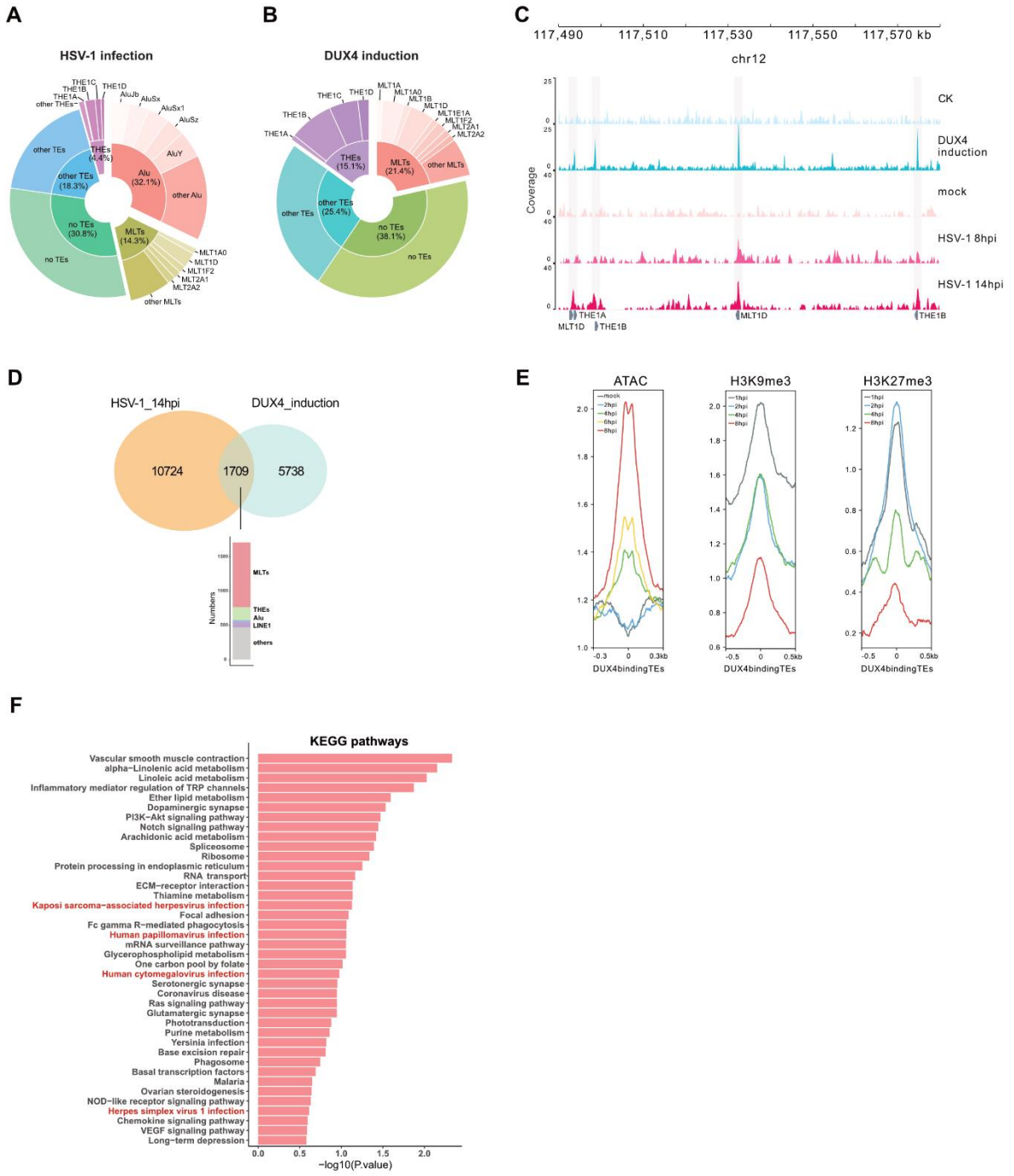

Figure 3: Exploring DUX4-Mediated TE Regulation in Viral Infections. (**A**) Distribution of DUX4 binding to TEs in ChIP-seq data of HSV-1-infected cells, categorized into THEs, MLTs, Alu, other TE binding, and no-TE binding. (**B**) Distribution of DUX4 binding to TEs in ChIP-seq data of DUX4-induced HFF cells, categorized into THEs, MLTs, other TEs binding, and no-TEs binding. (**C**) RNA-Seq. and DUX4 ChIP-seq tracks of TE binding in control (CK) and DUX4 induced cells, as well as cells infected with HSV-1 (mock, 8 hpi, and 14 hpi). (**D**) Venn diagram illustrating the overlap of TEs bound by DUX4 at 14h post HSV-1 infection in HFF cells and DUX4 binding after DUX4 induction in HFF cells. The overlapping TEs include MLT-, THE-, Alu-, and LINE1-elements. (**E**) Peak count frequency of ATAC-seq and ChIP-seq of transcriptional repression mark (H3K9me3 and H3K27me3) peaks overlapping with DUX4 binding at TEs. (**F**) KEGG pathways analysis of genes proximal to DUX4-binding sites within TEs (1 kb up and downstream).

Role of DUX4 in TEs activation

Given that approximately 70% of DUX4 binding sites are within TEs, our investigation aimed to elucidate the influence of DUX4 on TE activation. We reanalyzed RNA-seq datasets derived from wild-type- (WT) and DUX4 knockout- (KO) cells at various time points (4, 6, 8, and 10 h) before and after exposure to HSV-1(24). As anticipated, TEs activation surged from 6 hpi onward upon HSV-1 infection. However, upon infection of DUX4 KO cells, the expression of TEs was drastically reduced, although TE-activation was still detectable (Fig.4A,B). We observed that some activated TEs in WT cells are not regulated by DUX4 as shown by comparison of DUX4-wt and -ko cells (clusters 1, 3 and 12, Fig.4A). More than 75% of TEs demonstrated lower expression in DUX4 KO cells across all time points, with expression of TEs reduced about 99% at 10 hpi (Fig.4C). Although the KO cells have a significantly lower expression of viral transcripts at 8h and 10h post infection (Fig.S3B), the big difference in TE expression levels is unlikely to be a result of the reduced viral replication in the ko cells. Moreover, when we overexpressed DUX4 in 293T cells, over 105 TEs, including MLT-, THE-, LINE-, and Alu-elements were activated (Fig.4D). 97% of activated TEs in DUX4-induced cells overlapped with TEs down-regulated in DUX4 KO cells upon HSV-1 infection, demonstrating that expression of respective TEs is indeed dependent on DUX4 (Fig.S3A). These findings underscore the pivotal role of DUX4 in the regulation of TEs-expression upon HSV-1 infection.

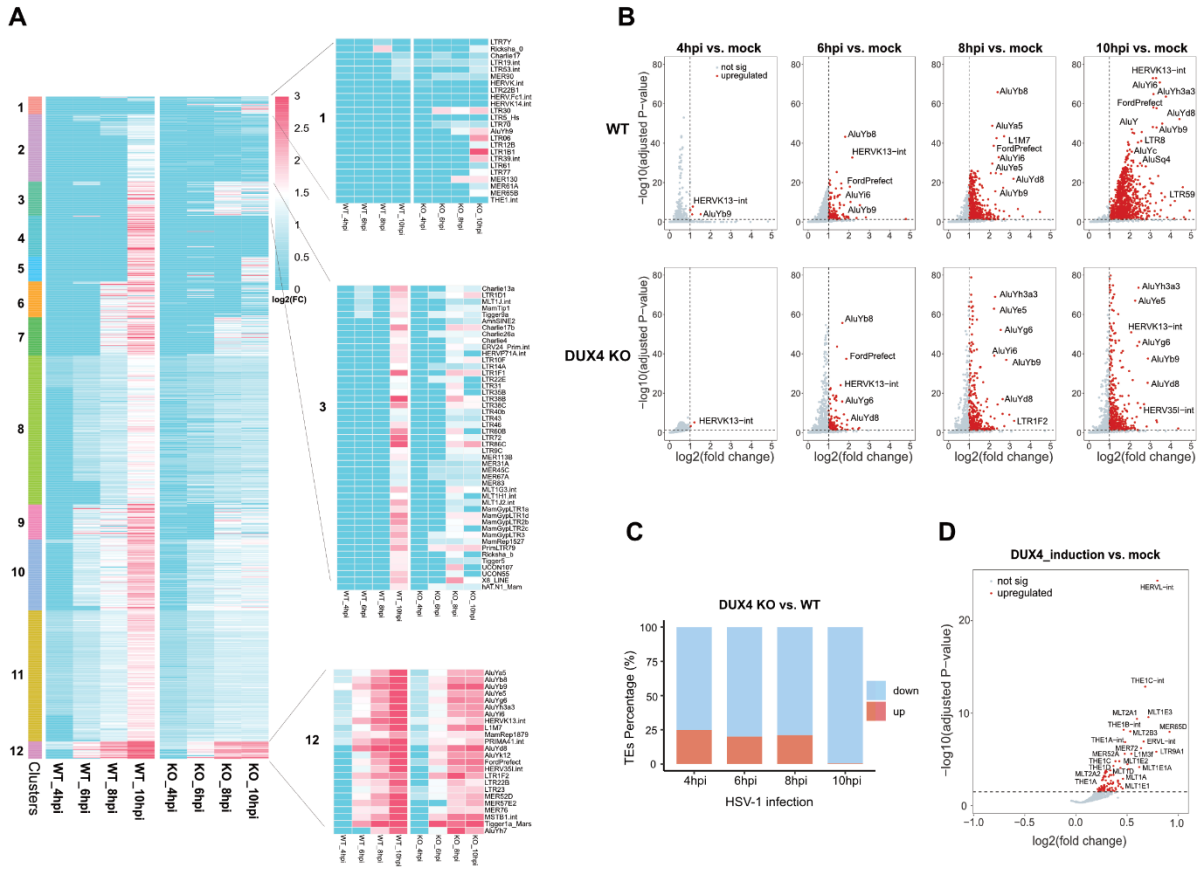

Figure 4: Differential TEs Regulation in Wild-Type (WT) and DUX4 Knockout (KO) Cells during HSV-1 Infection. (**A**) Heatmap depicting log2(FC) of upregulated TEs at 4, 6, 8, and 10 hours post HSV-1 infection (hpi) compared to mock in both WT and DUX4 KO HAP1 cells. TEs cluster in 12 clusters according to their expression kinetics. Values exceeding log2(FC) > 3 are capped at 3. Subclusters 1, 3, and 12 highlight TEs that are not regulated by DUX4. (**B**) Volcano plot illustrating significantly upregulated TEs upon HSV-1 infection at 4, 6, 8, and 10 hpi in WT and DUX4 KO. Red dots indicate TEs meeting the criteria of log2(FC) > 1 and P < 0.05. (**C**) Percentage of up and downregulated TEs in KO compared to WT during HSV-1 infection at 4, 6, 8, and 10 hpi. (**D**) Volcano plot showing significantly upregulated TEs when overexpressed DUX4. Red dots indicate TEs meeting the criteria of P < 0.05.

KAP1 controls DUX4- and TE-activation

KAP1 is known to be a master regulator of TEs and functions as a highly SUMOylated protein that orchestrates the assembly of the TEs silencing machinery. This is chiefly accomplished through its recruitment of the histone methyltransferase SETDB1 in a SUMO-dependent manner (48). Notably, prior studies have indicated that KAP1 can suppress the DUX4 homologue DUX in a LINE1-dependent fashion in mice (49). Our investigation unveils a novel aspect of KAP1 function, showing that KAP1 specifically

binds to the promoter region of DUX4 in humans (Fig.5A). Intriguingly, upon HSV-1 infection, KAP1 undergoes phosphorylation at Serine 473 and Serine 824 (Fig.5B). Phosphorylation of Serine 824 is dependent on ATM activation, as it can be blocked by the ATM-inhibitor KU60019, and is therefore part of the cellular response to viral infection. In contrast, Serine 473 phosphorylation is ATM independent and directly mediated by the HSV-1 kinase US3 (Fig.5C). Serine 473 phosphorylation is known to interfere with transcriptional repression of KRAB-zinc finger protein (KRAB-ZFP) target genes (50), thus disrupting its function and potentially triggering the activation of DUX4 and subsequent TEs. To test our hypothesis, we analyzed RNA-seq datasets derived from WT and KAP1 knockdown (KD) cells both before and after exposure to HCMV. Remarkably, KAP1 KD dramatically induces the expression of DUX4 target genes in the absence of infection (Fig.5D) as well as the induction of TEs (Fig.5E). Comparative analysis of KAP1 KD cells with and without HCMV infection reveals increased expression of TEs in the presence of virus (Fig.5F), underscoring the dependence of DNA virus-induced TEs activation on KAP1 and DUX4. Our results not only confirm the role of KAP1 in the regulation of TEs, but also establish a KAP1- and DUX4-dependent mechanism for herpesvirus-mediated TEs activation. Interestingly, when we examined the TEs upregulated during HSV-1 and HCMV infection, the majority were found to be KAP1-specific, with approximately 358 and 203 TEs regulated by both DUX4 and KAP1, respectively (Fig.5G,H). This multifaceted interplay between KAP1, DUX4 and DNA viruses sheds light on the intricate regulatory network governing TEs expression.

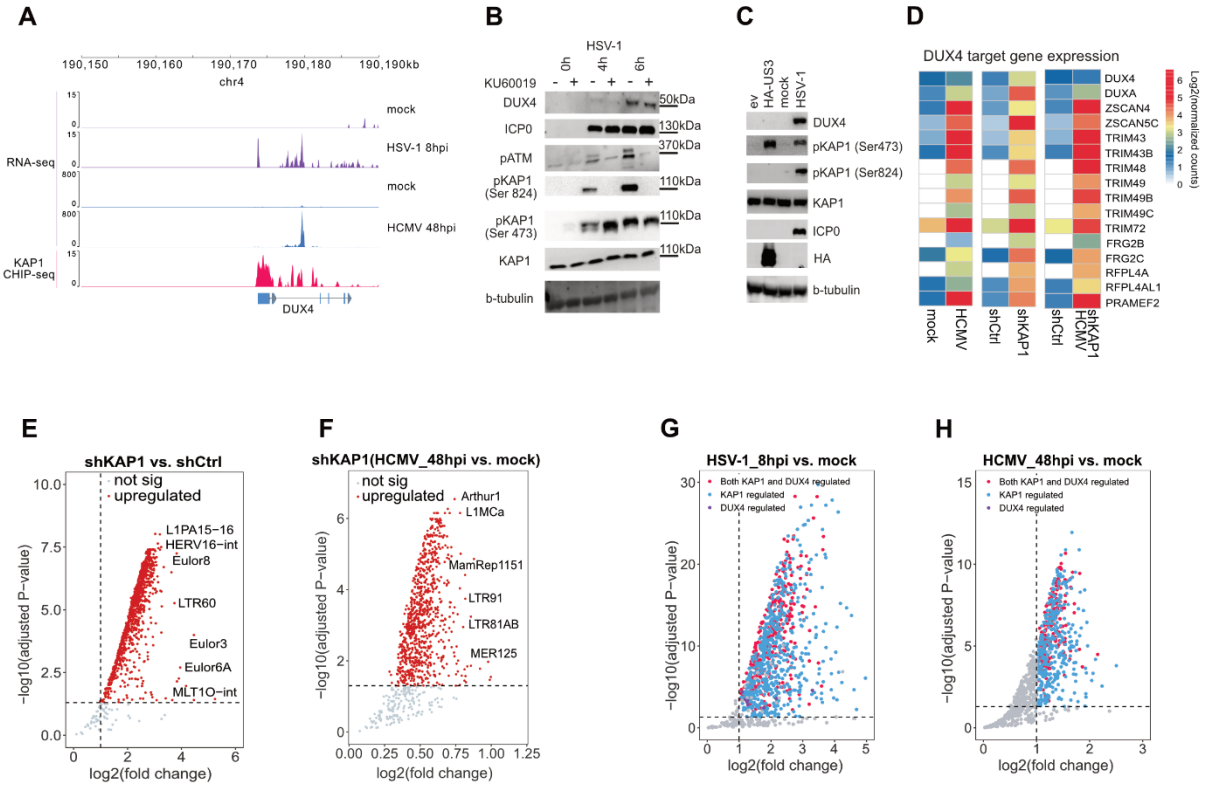

Figure 5: Investigating the Role of DUX4 and KAP1 in TEs activation upon Viral Infections. (**A**) Schematic representation and RNA-seq tracks illustrate the expression of DUX4 after HSV-1 and HCMV infection. Schematic representation and KAP1 ChIP-seq tracks showed KAP1 specifically binds to the promoter region of DUX4. (**B**) Western blot analysis of HFF cells infected with HSV-1 for 14h with MOI of 5. Including expression of DUX4 and phosphorylation of KAP1 and ATM. Tubulin serves as loading control. ICP0 serves as a control for viral infection. (**C**) Transfection of 293T cells with a plasmid encoding for HA-tagged HSV-1 kinase US3. 293T infected with HSV-1 served as control. Western Blot analysis of DUX4, KAP1, phospho-KAP1, ICP0 as infection control and tubulin as loading control. One representative experiment of n=3 is shown. (**D**) Heatmap displaying the number of normalized reads per gene (log2) with a focus on DUX4 and DUX4-target genes. Left to right panel: HFF cells infected with HCMV for 48h; HFF cells transfected with shCtrl and shKAP1; HFF cells transfected with shCtrl and shKAP1 and infected with HCMV for 48h. (**E**) Volcano plot depicting upregulated TEs in HFF cells transfected with shKAP1. Red dots represent significantly upregulated TEs from the TEs analysis (log2(FC)>1, $P < 0.05$). (**F**) Volcano plot illustrating upregulated TEs in HFF cells transfected with shKAP1 upon HCMV infection for 48 hpi. Red dots represent significantly upregulated TEs from the TEs analysis ($P < 0.05$). (**G**) Volcano plot showcasing upregulated TEs upon HSV-1 infection at 8 hpi. Red dots represent TEs significantly upregulated by both KAP1 and DUX4, blue dots represent TEs upregulated by KAP1, and purple dots represent TEs upregulated by DUX4 (log2(FC)>1, $P < 0.05$). (**H**) Volcano plot illustrating upregulated TEs upon HCMV infection at 48 hpi. Red dots represent TEs significantly upregulated by both KAP1 and DUX4, blue dots represent TEs upregulated by KAP1, and purple dots represent TEs upregulated by DUX4 (log2(FC)>1, $P < 0.05$).

TEs are upregulated in DNA virus associated human cancers

We recently showed that the germline transcription factor DUX4 is expressed in the context of DNA virus infection, namely herpesviruses (HSV-1, HSV-2, HCMV, KSHV and EBV), papillomaviruses (HPV16) and polyomaviruses (MCPyV) (51). Along this line, we wanted to investigate whether herpesviral infections also induce TEs expression in vivo. By re-analyzing single-cell RNA-sequencing data of tumor samples from EBV-positive nasopharyngeal carcinoma patients, we were able to show that TEs are upregulated in EBV-positive tumor cells (Fig.6A; Fig.S4A). We noted that MLT- and THE-family transposons expression is significantly higher in malignant cells expressing viral transcripts compared to bystander cells, indicating that TEs expression is driven by the virus infection (Fig.6A; Fig.S4A). Next, we investigated whether tumor cells of MCPyV-positive Merkel cell carcinoma patients and HPV-positive head and neck squamous cell carcinoma patients also have elevated TEs levels. Similarly to EBV-tumors, in Merkel cell carcinoma tumor patients, the expression of MLT and THE members is significantly upregulated in MCV-expressing cells compared to other cell types (Fig.6B; Fig.S4B,C). Moreover, in head and neck squamous cell carcinoma patients, the expression of these TEs is significantly upregulated in HPV-enriched malignant cells compared to bystander cells (Fig.6C; Fig.S4D). Taken together, these data suggest that TE-expression is a common feature of human DNA virus infection *in vivo* in the context of different types of tumors. Moreover, it is known that transposable elements serve as transcription factor binding sites that can activate downstream genes. Therefore, we analyzed splicing events from TEs to downstream exons of genes. Using conservative filtering criteria, we found eight examples of genes that are activated upon infection and DUX4 overexpression (two examples shown in Fig. 6D) due to DUX4 binding to a TE upstream of a gene.

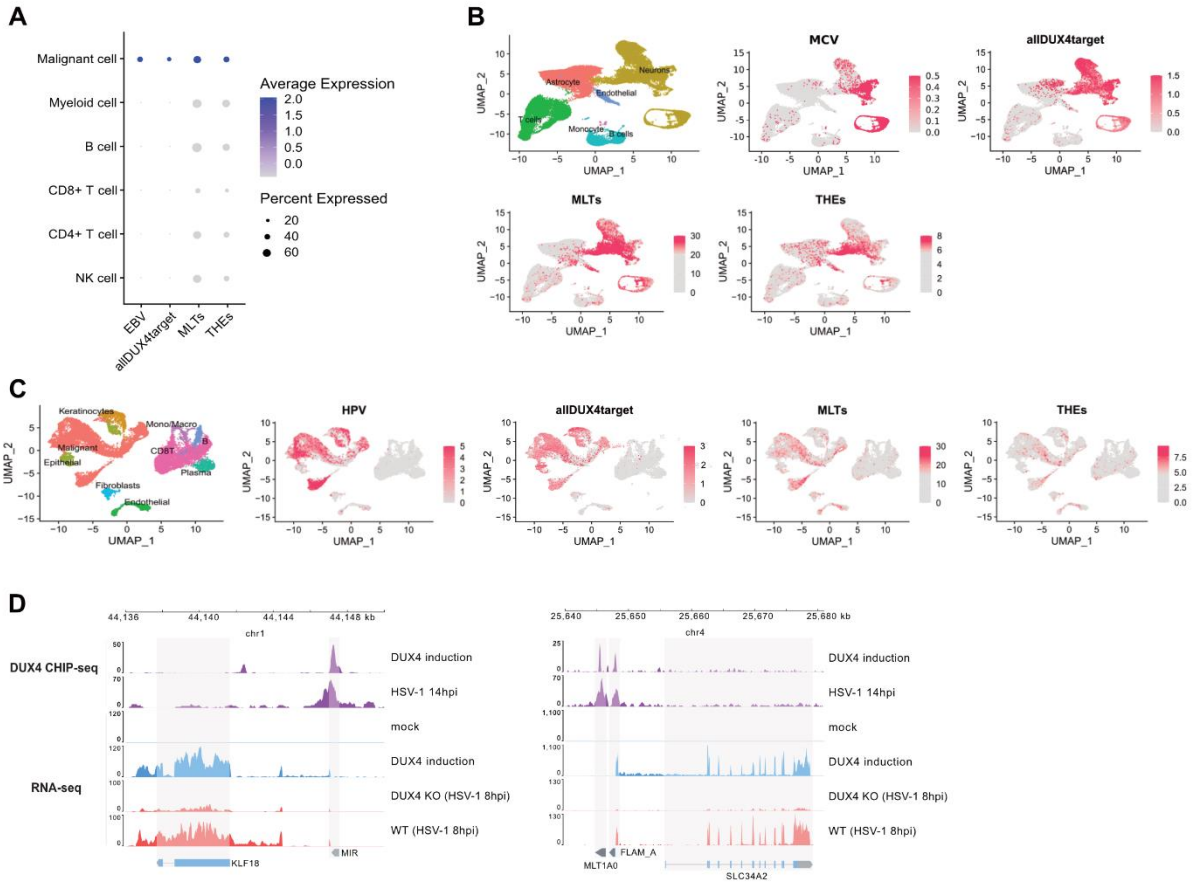

Figure 6: Single-Cell RNA-seq Analysis of TEs in Virally Associated Tumors and the examples of TEs work as alternative promoters. (**A**) Nasopharyngeal Carcinoma Patients with Epstein-Barr Virus (EBV) Positive: Dotplot showcasing the number of normalized EBV-specific reads per cell, normalized DUX4-target reads per cell, normalized MLTs reads per cell, and normalized THEs reads per cell in different cell types. (**B**) MCC Tumor Patients with Merkel Cell Polyomavirus (MCV) Positive: First panel: Cell identity map. Second panel: UMAP projection of normalized MCV-specific reads per cell. Third panel: UMAP projection of normalized DUX4-target reads per cell. Fourth panel: UMAP projection of normalized MLTs reads per cell. Fifth panel: UMAP projection of normalized THEs reads per cell. (**C**) Head and Neck Squamous Cell Carcinoma (HNSCC) Patients with Human Papillomavirus (HPV) Positive: First panel: Cell identity map. Second panel: UMAP projection of normalized HPV-specific reads per cell. Third panel: UMAP projection of normalized DUX4-target reads per cell. Fourth panel: UMAP projection of normalized MLTs reads per cell. Fifth panel: UMAP projection of normalized THEs reads per cell. (**D**) Left panel: schematic representation and CHIP-seq tracks illustrate the binding of DUX4 to the TEs MIR that is located in the promoter region of the gene KLF18 after DUX4 induction and HSV-1 infection. Right panel: schematic representation and RNA-seq tracks illustrate the expression of the new transcript of gene KLF18 after DUX4 induction and HSV-1 infection. Schematic representation and CHIP-seq tracks illustrate the binding of DUX4 to the TEs MLT1A0 and FLAM_A that are located in the promoter region of the gene SLC34A2 after DUX4 induction and HSV-1 infection as well as RNA-seq. tracks of the new transcript of gene SLC34A2 after DUX4 induction and HSV-1 infection.

Discussion

The relationship between hosts and viruses involves ongoing co-evolution, with viruses exerting persistent selection pressure that drives the evolution of antiviral mechanisms within host cells. Our comprehensive transcriptomic analysis of human cells infected with HSV-1 and HCMV reveals a noteworthy upregulation of TEs post infection. This observation is further validated through single-cell RNA-seq analysis of patients with MCV, HPV, and EBV-associated cancers in vivo, emphasizing the clinical relevance of these findings. Despite previous reports indicating that viral infections can impact TE expression (48), the exact molecular mechanisms underlying this widespread impact remained elusive until now. A pivotal discovery of our study is the identification of the germline transcription factor DUX4 as a key player in the activation of TEs induced by DNA viruses. Utilizing DUX4 ChIP-seq, we found a significant association, with 69% of DUX4 binding occurring at TEs. The functional relevance of DUX4 is underscored by experiments demonstrating that its overexpression significantly induces TEs expression, while its knockout diminishes TEs expression upon HSV-1 infection. In addition, we found a strong association of DUX4 binding with opening of chromatin at DUX4 binding sites, indicated by ATAC-Seq. peaks and reduction of H3K9-trimethylation. This positions DUX4 as a central regulator, orchestrating the cascade leading to TEs activation during DNA viral infections.

In prior research, the role of SUMOylation-modified KAP1 in TEs-silencing has been demonstrated, triggering the establishment of repressive histone marks at TEs sites (52). Notably, the SUMOylated form of KAP1 exhibits rapid disappearance following Influenza A virus (IAV) infection (18)(53), a phenomenon proposed to result from the phosphorylation of KAP1 (54). Phosphorylation of KAP1 is known to regulate binding to heterochromatin protein 1 (HP1), KRAB-ZFP and the histone methyltransferase SETDB1. Our investigation extends these findings by demonstrating the

phosphorylation of KAP1 during HSV-1 infection at two different residues, one mediated by ATM-signaling and the second one mediated by the viral kinase US3. The ChIP-seq analysis discloses binding of KAP1 to the DUX4 promoter and KAP1 knockdown induced DUX4 target genes and TEs expression, suggesting its regulatory role in modulating DUX4 and subsequently influencing the expression of TEs. This discovery illuminates a novel pathway by which DNA viruses, exemplified by HSV-1, may induce the activation of TEs through the phosphorylation of KAP1 and subsequent modulation of DUX4.

The interplay between viruses and their hosts includes the reactivation and co-optation of TEs for effective antiviral defense, as highlighted in previous studies (48). In general, it is thought that the activation of TEs results in the activation of innate immune pathways because TEs are derived from viruses and therefore still contain pathogen associated molecular patterns (PAMPs) that might be recognized by pattern recognition receptors (PRRs). Activation of PRRs results in triggering cellular defense mechanisms e.g. expression of interferon stimulated genes and production of interferons that act on neighboring cells and activate antiviral defense mechanisms. However, we showed that activation of DUX4 is directly mediated by herpesviral proteins; for HSV-1 two immediate-early proteins, ICP0 and ICP4 are sufficient to induce expression of DUX4 (24). In addition, the HSV-1 kinase US3 phosphorylates KAP1 at Ser473, thereby lifts restriction by KAP1, which results in TEs expression. This would argue against an antiviral mechanism since viruses generally avoid activation of antiviral mechanisms by all means. In addition, we previously showed that DUX4 activation is critical for efficient replication of HSV-1 and HSV-2 (24). Therefore, it might be conceivable that activation of TEs has beneficial effects for the replication or persistence of DNA viruses. Members of herpesviruses, papillomaviruses and polyomaviruses are known to establish persistent infections that lead to oncogenic

transformation of cells in a proportion of cases, and we could demonstrate that TEs are activated in cancer cells containing viral transcripts from patients (Fig.6A,B,C). In this context, our present investigation significantly contributes to the understanding of the complex relationship between DNA viruses and the host genome in respect to TE-activation.

DNA virus genomes contain multiple genes that are derived from host genes. For herpesviruses, some of the genes are conserved across all human herpesviruses, like the large ribonucleotide reductase subunit, whereas others are only present in individual subfamilies like the viral FGARATs of gamma-herpesviruses KSHV and EBV (55). It is thought that the genes were “stolen” by viruses at one point during evolution by horizontal gene transfer and then diversified during the coevolution with the respective hosts. The viral FGARATs for example show about 25% sequence identity with human FGARAT and have lost their role in nucleotide metabolism (56). However, viral FGARATs have evolved as potent inhibitors of the antiviral function of PML nuclear bodies (57, 58). Host gene homologues are also present in the genome of Poxviruses. Poxviruses are also DNA viruses, but in contrast to herpesviruses, they replicate in the cytoplasm. Two recent papers showed that the uptake of host genes into the poxvirus genome is mediated by retrotransposons (59, 60). The authors showed that mRNA can be reverse transcribed by the reverse transcriptase of LINE-1 elements and then integrate into poxvirus genomes. Although the exact mechanism might be different between nuclear and cytoplasmic replicating viruses due to different replication strategies, it shows an important role for retrotransposons in the host adaptation of DNA viruses. It is conceivable that the induction of retrotransposons is crucial for the evolution of DNA viruses and subsequent studies will be needed to investigate which role TEs activation plays in DNA virus infection and viral oncogenesis in detail.

The results presented here not only provide a characterization of TEs activation during DNA virus infection but also highlight the essential roles of DUX4 and KAP1 as key regulators in this process. The intricate molecular mechanisms uncovered in this study offer valuable insights into the dynamics of host genome-virus interactions. Understanding the role and regulation of TEs during viral infections is crucial, as TEs are integral components of cis-regulatory elements, and their dysregulation can have profound implications for host genomic stability and function. These findings have implications for broader fields, including virology and cancer biology, as the study includes analysis of viral-associated cancers. The identification of DUX4 as a key mediator in TEs activation suggests potential therapeutic targets for controlling TEs expression during viral infections, with potential implications for diseases associated with viral etiology. Overall, this study significantly advances our knowledge of the molecular mechanisms governing TEs activation in the context of DNA virus infections, opening avenues for further research and therapeutic exploration.

Data Availability

The raw sequencing datasets have been uploaded to the Gene Expression Omnibus (GEO) database, including RNA-Seq of DUX4 over-expression (GSE253543) and RNA-Seq of HCMV infection and KAP1 knockdown (GSE256111). The study used publicly available datasets from GEO, including bulk RNA-seq of HSV-1 infection (GSE59717) (61), ATAC-seq of HSV-1 infection (GSE100611) (62), CHIP-seq of H3K9me3 and H3K27me3 upon HSV-1 infection (GSE124803) (47), CHIP-seq of KAP1 (GSE72622) (63), RNA-seq of DUX4-KO (GSE234489) (51), CHIP-seq of DUX4 (GSM837613 and GSE174759) (64)(51), as well as the single-cell RNA-seq data of nasopharyngeal carcinoma (GSE162025) (65), Head and neck squamous cell carcinoma (GSE164690) (66) and Merkel cell carcinoma tumor (GSE226438) (67).

UCSC genome browser session links for the ATAC-seq. and ChIP-seq. data are available under the following link: https://genome-euro.ucsc.edu/s/tanjiang/TEs_track.

Acknowledgements

We are grateful to Dr. Daniel Sauter (Tübingen, Germany) and Dr. Andrea Thoma-Kress (Erlangen, Germany) for providing plasmids. This study was supported by grants from the German Research Foundation (DFG En 423/5-1, to A.E.), the BMBF Junior Research Group "Duxdrugs" (BMBF 01KI2017, F.F.) and the Deutsche Forschungsgemeinschaft in the framework of Research Unit FOR5200 DEEP-DV (443644894) through project STA357/8-1 to T.S.. E.N. is a member of the Spemann graduate school of biology and medicine (SGBM).

Declaration of interests

The authors declare no competing interests.

1. Klein,S.J. and O'Neill,R.J. (2018) Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res.*, **26**, 5–23.

2. Cordaux,R. and Batzer,M.A. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.*, **10**, 691–703.

3. Liang,Y., Qu,X., Shah,N.M. and Wang,T. (2024) Towards targeting transposable elements for cancer therapy. *Nat. Rev. Cancer*.

4. Wicker,T., Sabot,F., Hua-Van,A., Bennetzen,J.L., Capy,P., Chalhoub,B., Flavell,A., Leroy,P., Morgante,M., Panaud,O., *et al.* (2007) A unified classification system for

eukaryotic transposable elements. *Nat. Rev. Genet.*, **8**, 973–982.

5. Dupressoir,A., Vernochet,C., Bawa,O., Harper,F., Pierron,G., Opolon,P. and Heidmann,T. (2009) Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 12127–12132.

6. Torre,D., Francoeur,N.J., Kalma,Y., Gross Carmel,I., Melo,B.S., Deikus,G., Allette,K., Flohr,R., Fridrikh,M., Vlachos,K., *et al.* (2023) Isoform-resolved transcriptome of the human preimplantation embryo. *Nat. Commun.*, **14**, 1–23.

7. Pastuzyn,E.D., Day,C.E., Kearns,R.B., Kyrke-Smith,M., Taibi,A.V., McCormick,J., Yoder,N., Belnap,D.M., Erlendsson,S., Morado,D.R., *et al.* (2018) The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell*, **173**, 275.

8. Choi,Y.J., Lin,C.-P., Risso,D., Chen,S., Kim,T.A., Tan,M.H., Li,J.B., Wu,Y., Chen,C., Xuan,Z., *et al.* (2017) Deficiency of microRNA expands cell fate potential in pluripotent stem cells. *Science*, **355**.

9. Hackett,J.A., Kobayashi,T., Dietmann,S. and Surani,M.A. (2017) Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. *Stem Cell Reports*, **8**, 1645–1658.

10. Miao,B., Fu,S., Lyu,C., Gontarz,P., Wang,T. and Zhang,B. (2020) Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol.*, **21**, 255.

11. Sundaram,V., Cheng,Y., Ma,Z., Li,D., Xing,X., Edge,P., Snyder,M.P. and Wang,T. (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.*, **24**, 1963–1976.

12. Boroviak,T., Stirparo,G.G., Dietmann,S., Hernando-Herraez,I., Mohammed,H., Reik,W., Smith,A., Sasaki,E., Nichols,J. and Bertone,P. (2018) Single cell

transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development*, **145**.

13. Pasquesi,G.I.M., Perry,B.W., Vandewege,M.W., Ruggiero,R.P., Schield,D.R. and Castoe,T.A. (2020) Vertebrate Lineages Exhibit Diverse Patterns of Transposable Element Regulation and Expression across Tissues. *Genome Biol. Evol.*, **12**, 506–521.

14. Frank,J.A., Singh,M., Cullen,H.B., Kirou,R.A., Benkaddour-Boumzaouad,M., Cortes,J.L., Garcia Pérez,J., Coyne,C.B. and Feschotte,C. (2022) Evolution and antiviral activity of a human protein of retroviral origin. *Science*, **378**, 422–428.

15. Bravo,J.I., Nozownik,S., Danthi,P.S. and Benayoun,B.A. (2020) Transposable elements, circular RNAs and mitochondrial transcription in age-related genomic regulation. *Development*, **147**.

16. Jang,H.S., Shah,N.M., Du,A.Y., Dailey,Z.Z., Pehrsson,E.C., Godoy,P.M., Zhang,D., Li,D., Xing,X., Kim,S., *et al.* (2019) Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.*, **51**, 611–617.

17. Shah,N.M., Jang,H.J., Liang,Y., Maeng,J.H., Tzeng,S.-C., Wu,A., Basri,N.L., Qu,X., Fan,C., Li,A., *et al.* (2023) Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nat. Genet.*, **55**, 631–639.

18. Schmidt,N., Domingues,P., Golebiowski,F., Patzina,C., Tatham,M.H., Hay,R.T. and Hale,B.G. (2019) An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 17399–17408.

19. Neugebauer,E., Bastidas-Quintero,A.M., Weidl,D. and Full,F. (2023) Pioneer factors in viral infection. *Front. Immunol.*, **14**, 1286617.

20. Sorek,M., Meshorer,E. and Schlesinger,S. (2022) Impaired activation of transposable elements in SARS-CoV-2 infection. *EMBO Rep.*, **23**, e55101.

21. Chen, X., Pacis, A., Aracena, K. A., Gona, S., Kwan, T., Groza, C., Lin, Y.L.,

Sindeaux R., Yotova V., Pramatarova, A., Simon, MM., Pastinen, T., Barreiro, LB., Bourque, G. (2023) Transposable elements are associated with the variable response to influenza infection. *Cell Genomics*, **3**, 100292.

22. Srinivasachar Badarinarayan,S., Shcherbakova,I., Langer,S., Koepke,L., Preising,A., Hotter,D., Kirchhoff,F., Sparrer,K.M.J., Schotta,G. and Sauter,D. (2020) HIV-1 infection activates endogenous retroviral promoters regulating antiviral gene expression. *Nucleic Acids Res.*, **48**, 10890–10908.

23. Genome graphs detect human polymorphisms in active epigenomic state during influenza infection (2023) *Cell Genomics*, **3**, 100294.

24. Full,F., Walter,S., Neugebauer,E., Tan,J., Drayman,N., Franke,V., Tay,S., Landthaler,M., Akalin,A., Ensser,A., *et al.* (2023) Herpesviruses mimic zygotic genome activation to promote viral replication. *Res Sq*, 10.21203/rs.3.rs-3125635/v1.

25. Yang,S.-H., Jaffray,E., Hay,R.T. and Sharrocks,A.D. (2003) Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. *Mol. Cell*, **12**, 63–74.

26. Zufferey,R., Nagy,D., Mandel,R.J., Naldini,L. and Trono,D. (1997) Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat. Biotechnol.*, **15**, 871–875.

27. Yurkovetskiy,L., Guney,M.H., Kim,K., Goh,S.L., McCauley,S., Dauphin,A., Diehl,W.E. and Luban,J. (2018) Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. *Nat Microbiol*, **3**, 1354–1361.

28. Sinzger,C., Hahn,G., Digel,M., Katona,R., Sampaio,K.L., Messerle,M., Hengel,H., Koszinowski,U., Brune,W. and Adler,B. (2008) Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J. Gen. Virol.*, **89**, 359–368.

29. Scherer,M., Otto,V., Stump,J.D., Klingl,S., Müller,R., Reuter,N., Muller,Y.A., Sticht,H. and Stamminger,T. (2016) Characterization of Recombinant Human Cytomegaloviruses Encoding IE1 Mutants L174P and 1-382 Reveals that Viral Targeting of PML Bodies Perturbs both Intrinsic and Innate Immune Responses. *J. Virol.*, **90**, 1190–1205.

30. Bhardwaj,V., Heyne,S., Sikora,K., Rabbani,L., Rauer,M., Kilpert,F., Richter,A.S., Ryan,D.P. and Manke,T. (2019) snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics*, **35**, 4757–4759.

31. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2012) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

32. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

33. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–5.

34. 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**, 550.

35. Jin,Y., Tam,O.H., Paniagua,E. and Hammell,M. (2015) TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, **31**, 3593–3599.

36. Kaminow,B., Yunusov,D. and Dobin,A. (2021) STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv*, 10.1101/2021.05.05.442755.

37. Hao,Y., Stuart,T., Kowalski,M.H., Choudhary,S., Hoffman,P., Hartman,A.,

Srivastava,A., Molla,G., Madad,S., Fernandez-Granda,C., *et al.* (2023) Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.*

38. He,J., Babarinde,I.A., Sun,L., Xu,S., Chen,R., Shi,J., Wei,Y., Li,Y., Ma,G., Zhuang,Q., *et al.* (2021) Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat. Commun.*, **12**, 1–14.

39. Aran,D., Looney,A.P., Liu,L., Wu,E., Fong,V., Hsu,A., Chak,S., Naikawadi,R.P., Wolters,P.J., Abate,A.R., *et al.* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.*, **20**, 163–172.

40. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

41. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.

42. Lun,A.T.L. and Smyth,G.K. (2016) csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.*, **44**, e45.

43. Chen,H. and Boutros,P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, **12**, 1–7.

44. Wickham,H. (2009) ggplot2: Elegant Graphics for Data Analysis Springer Science & Business Media.

45. Lopez-Delisle,L., Rabbani,L., Wolff,J., Bhardwaj,V., Backofen,R., Grüning,B., Ramírez,F. and Manke,T. (2021) pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics*, **37**, 422–423.

46. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.

47. The epigenetic landscapes of histone modifications on HSV-1 genome in human

THP-1 cells (2020) *Antiviral Res.*, **176**, 104730.

48. Hale,B.G. (2022) Antiviral immunity triggered by infection-induced host transposable elements. *Curr. Opin. Virol.*, **52**, 211–216.

49. Percharde,M., Lin,C.-J., Yin,Y., Guan,J., Peixoto,G.A., Bulut-Karslioglu,A., Biechele,S., Huang,B., Shen,X. and Ramalho-Santos,M. (2018) A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*, **174**, 391–405.e19.

50. Hu,C., Zhang,S., Gao,X., Gao,X., Xu,X., Lv,Y., Zhang,Y., Zhu,Z., Zhang,C., Li,Q., *et al.* (2012) Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1 Ser-473 Phosphorylation in DNA Damage Response. *J. Biol. Chem.*, **287**, 18937–18952.

51. Walter,S., Franke,V., Drayman,N., Wyler,E., Tay,S., Landthaler,M., Akalin,A., Ensser,A. and Full,F. (2021) Herpesviral induction of germline transcription factor DUX4 is critical for viral gene expression. *bioRxiv*, 10.1101/2021.03.24.436599.

52. Geis,F.K. and Goff,S.P. (2020) Silencing and Transcriptional Regulation of Endogenous Retroviruses: An Overview. *Viruses*, **12**.

53. Domingues,P., Golebiowski,F., Tatham,M.H., Lopes,A.M., Taggart,A., Hay,R.T. and Hale,B.G. (2015) Global Reprogramming of Host SUMOylation during Influenza Virus Infection. *Cell Rep.*, **13**, 1467–1480.

54. Krischuns,T., Günl,F., Henschel,L., Binder,M., Willemsen,J., Schloer,S., Rescher,U., Gerlt,V., Zimmer,G., Nordhoff,C., *et al.* (2018) Phosphorylation of TRIM28 Enhances the Expression of IFN-β and Proinflammatory Cytokines During HPAIV Infection of Human Lung Epithelial Cells. *Front. Immunol.*, **9**, 2229.

55. Holzerlandt,R., Orengo,C., Kellam,P. and Albà,M.M. (2002) Identification of new herpesvirus gene homologs in the human genome. *Genome Res.*, **12**, 1739–1748.

56. Full,F., Reuter,N., Zielke,K., Stamminger,T. and Ensser,A. (2012) Herpesvirus Saimiri Antagonizes Nuclear Domain 10-Instituted Intrinsic Immunity via an ORF3-

Mediated Selective Degradation of Cellular Protein Sp100. *J. Virol.*, 10.1128/jvi.06992-11.

57. Full,F., Jungnickl,D., Reuter,N., Bogner,E., Brulois,K., Scholz,B., Stürzl,M., Myoung,J., Jung,J.U., Stamminger,T., *et al.* (2014) Kaposi's Sarcoma Associated Herpesvirus Tegument Protein ORF75 Is Essential for Viral Lytic Replication and Plays a Critical Role in the Antagonization of ND10-Instituted Intrinsic Immunity. *PLoS Pathog.*, **10**, e1003863.

58. Tsai,K., Thikmyanova,N., Wojcechowskyj,J.A., Delecluse,H.-J. and Lieberman,P.M. (2011) EBV tegument protein BNRF1 disrupts DAXX-ATRX to activate viral early gene transcription. *PLoS Pathog.*, **7**, e1002376.

59. Rahman,M.J., Haller,S.L., Stoian,A.M.M., Li,J., Brennan,G. and Rothenburg,S. (2022) LINE-1 retrotransposons facilitate horizontal gene transfer into poxviruses. *Elife*, **11**.

60. Fixsen,S.M., Cone,K.R., Goldstein,S.A., Sasani,T.A., Quinlan,A.R., Rothenburg,S. and Elde,N.C. (2022) Poxviruses capture host genes by LINE-1 retrotransposition. *Elife*, **11**.

61. Rutkowski,A.J., Erhard,F., L'Hernault,A., Bonfert,T., Schilhabel,M., Crump,C., Rosenstiel,P., Efstathiou,S., Zimmer,R., Friedel,C.C., *et al.* (2015) Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.*, **6**, 1–15.

62. Hennig,T., Michalski,M., Rutkowski,A.J., Djakovic,L., Whisnant,A.W., Friedl,M.-S., Jha,B.A., Baptista,M.A.P., L'Hernault,A., Erhard,F., *et al.* (2018) HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLoS Pathog.*, **14**, e1006954.

63. McNamara,R.P., Reeder,J.E., McMillan,E.A., Bacon,C.W., McCann,J.L. and D'Orso,I. (2016) KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables

Transcription Elongation by RNA Polymerase II. *Mol. Cell*, **61**, 39–53.

64. Geng,L.N., Yao,Z., Snider,L., Fong,A.P., Cech,J.N., Young,J.M., van der Maarel,S.M., Ruzzo,W.L., Gentleman,R.C., Tawil,R., *et al.* (2012) DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell*, **22**, 38–51.

65. Liu,Y., He,S., Wang,X.-L., Peng,W., Chen,Q.-Y., Chi,D.-M., Chen,J.-R., Han,B.-W., Lin,G.-W., Li,Y.-Q., *et al.* (2021) Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat. Commun.*, **12**, 1–18.

66. Kürten,C.H.L., Kulkarni,A., Cillo,A.R., Santos,P.M., Roble,A.K., Onkar,S., Reeder,C., Lang,S., Chen,X., Duvvuri,U., *et al.* (2021) Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat. Commun.*, **12**, 1–16.

67. Das,B.K., Kannan,A., Velasco,G.J., Kunika,M.D., Lambrecht,N., Nguyen,Q., Zhao,H., Wu,J. and Gao,L. (2023) Single-cell dissection of Merkel cell carcinoma heterogeneity unveils transcriptomic plasticity and therapeutic vulnerabilities. *Cell Rep Med*, **4**, 101101.