

# Exploring the transcriptomic profile of human monkeypox virus via CAGE and native RNA sequencing approaches

Gergely Ármin Nagy,<sup>1</sup> Dóra Tombácz,<sup>1</sup> István Prazsák,<sup>1</sup> Zsolt Csabai,<sup>1</sup> Ákos Dörmő,<sup>1</sup> Gábor Gulyás,<sup>1</sup> Gábor Kemenesi,<sup>2,3</sup> Gábor E. Tóth,<sup>2,3,4</sup> Jiří Holoubek,<sup>5,6,7</sup> Daniel Růžek,<sup>5,6,7</sup> Balázs Kakuk,<sup>1</sup> Zsolt Boldogkői<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 18.

**ABSTRACT** In this study, we employed short- and long-read sequencing technologies to delineate the transcriptional architecture of the human monkeypox virus and to identify key regulatory elements that govern its gene expression. Specifically, we conducted a transcriptomic analysis to annotate the transcription start sites (TSSs) and transcription end sites (TESs) of the virus by utilizing Cap Analysis of gene expression sequencing on the Illumina platform and direct RNA sequencing on the Oxford Nanopore technology device. Our investigations uncovered significant complexity in the use of alternative TSSs and TESs in viral genes. In this research, we also detected the promoter elements and poly(A) signals associated with the viral genes. Additionally, we identified novel genes in both the left and right variable regions of the viral genome.

**IMPORTANCE** Generally, gaining insight into how the transcription of a virus is regulated offers insights into the key mechanisms that control its life cycle. The recent outbreak of the human monkeypox virus has underscored the necessity of understanding the basic biology of its causative agent. Our results are pivotal for constructing a comprehensive transcriptomic atlas of the human monkeypox virus, providing valuable resources for future studies.

**KEYWORDS** monkeypox virus, CAGE-Seq, nanopore sequencing, long-read sequencing, transcriptome, poxvirus

Orthopoxvirus, a genus in the Poxviridae family, encompasses several significant human and animal pathogens. Orthopoxviruses include several species, most notably the variola virus, which causes smallpox, the monkeypox virus (MPXV), the cowpox virus, and the vaccinia virus (VACV), which is known for its use in smallpox vaccination (1–3). Over the course of centuries, smallpox claimed millions of lives until its successful eradication, thanks to an extensive worldwide vaccination initiative (4). MPXV is a zoonotic virus, originally endemic to Africa, that can cause human disease known as mpox (5). Typically, the symptoms are typically mild (6). The first human case was identified in the Democratic Republic of Congo (DRC) in 1970 (7). Since then, sporadic outbreaks have been increasingly reported in Eastern, Central, and Western Africa (8). Human MPXV (hMPXV) is classified into three phylogenetically distinct clades: clade I, the most virulent, with up to 10% human mortality, primarily transmitted by rodents in the Congo Basin (9); clade IIa, originally a zoonosis with low mortality in West Africa, which evolved into human-to-human transmission, causing an outbreak in Nigeria in 2017–2018 (9, 10); and clade IIb, responsible for the 2022 global outbreak, spreading through human-to-human transmission (9, 11, 12). The existence of multiple subclades indicates rapid evolution of these human-adapted lineages (9, 11, 12).

Orthopoxviruses have a large linear double-stranded DNA genome, approximately 200 kilobase pairs long (13). Unlike the majority of mammalian DNA viruses, including

**Editor** Michael J. Imperiale, University of Michigan, Ann Arbor, Michigan, USA

Address correspondence to Zsolt Boldogkői, boldogkoi.zsolt@med.u-szeged.hu.

Gergely Ármin Nagy and Dóra Tombácz contributed equally to this article. Author order was determined based on alphabetical order.

The authors declare no conflict of interest.

See the funding table on p. 18.

**Received** 26 April 2024

**Accepted** 31 July 2024

**Published** 27 August 2024

Copyright © 2024 Nagy et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

herpesviruses and adenoviruses, which replicate in the nucleus, poxviruses, along with the African swine fever virus, replicate in the cytoplasm. The replication and transcription processes of poxviruses are carried out within specialized structures known as “viral factories” (14). The regulation of viral gene expression is governed by transcription factors specific to different stages, which bind selectively to the promoters of early (E), intermediate (I), and late (L) genes (15). The full transcription machinery is pre-packaged within the poxvirus virion, which allows for immediate expression of E genes once the virus has entered the cell and while the viral genome is still encapsulated. This is then followed by DNA replication and the subsequent expression of I and L gene classes, collectively termed as post-replicative (PR) genes. E genes are responsible for encoding proteins that synthesize DNA and RNA molecules and those that play a part in the interactions between the virus and the host. Meanwhile, PR genes primarily encode the structural elements of the virus (16).

Unlike herpesviruses, which tend to produce 3'-co-terminal transcripts by the adjacent tandem genes, poxviruses generate a vast diversity of 3'-ends (17), especially during the late stages of infection (18). The lack of splicing in poxvirus transcripts is attributed to their replication in the cytoplasm (19). Poxviruses have the unique ability to produce their own enzymes for capping, decapping, and polyadenylation, and they employ strategies such as mRNA decapping to inhibit host translation (20). Though poxvirus mRNAs generally resemble host mRNAs in structure, one distinctive trait is the presence of 5'-poly(A) leaders in PR mRNAs (21). Recent studies have revealed that poly(A) leaders provide the capability to utilize either Cap-dependent or Cap-independent translation initiation (22).

Several studies have explored the transcriptional impact of hMPXV infection across various cell types, predominantly utilizing micro-array-based techniques (23–26). These pioneering works have laid the foundation for the understanding of the viral transcription landscape. A notable limitation of the micro-array-based techniques is their inability to resolve important aspects of the transcriptome, particularly to detect the transcript isoforms (27). Determining the exact genomic location of the transcription start sites (TSSs) and transcription end sites (TESs) of the mRNAs is crucial in annotating viral genomes. Methods, such as S1 nuclease treatment with labeled probe-hybridization (28, 29), have been developed as early attempts to determine both 5'-ends (30–34) and 3'-ends (35–37) of poxviral mRNAs. Rapid amplification of cDNA ends (RACE) (38) is a widely used PCR-based method to identify ends of cDNA transcripts. This technique was used to determine transcript boundaries in poxviruses (39). Poxviruses are unique among the viruses, because they have their own capping (40) and decapping enzymes (41, 42). Detection of Cap is utilized in transcriptome research for identifying transcription initiation (43, 44). Although microarray and PCR-based techniques offer high precision, they are limited to analyzing only those transcripts for which probes or gene-specific primers exist. In contrast, total RNA-sequencing methods allow for the examination of the entire transcriptome. With the advent of next-generation sequencers, the bulk analysis of whole transcriptome features, including TSSs and TESs, became possible.

Various sequencing methods are available to infer the whole transcriptome. Short-read sequencing (SRS) provides a high-throughput, base precision map of transcriptional activity. However, reverse transcription-dependent techniques are unable to circumvent the drawbacks occurring during cDNA synthesis, such as template switching (45, 46), false priming (47), or spurious antisense transcription (48).

Long-read sequencing (LRS) methods, such as single-molecule real-time and nanopore sequencing (49, 50), are able to read the entire mRNAs, making them indispensable in transcriptome research (51, 52). Oxford Nanopore Technologies (ONT) allows for the direct sequencing of native RNA molecules (dRNA-Seq). This approach eliminates the generation of false products that may arise during the library preparation process, specifically during the reverse transcription, second-strand synthesis, and PCR steps. The limitation of this technique is its reduced precision in annotating the 5'-ends of mRNAs (53). This issue can be mitigated by integrating dRNA-Seq with 5'-end-sensitive

PCR-free direct cDNA (dcDNA) sequencing (dcDNA-Seq) or selective detection of capped RNA ends by Cap Analysis of Gene Expression sequencing (CAGE-Seq) methods (54–56). The LRS cDNA-Seq approach has been applied for the analysis of dynamic VACV transcriptome (18, 27, 57, 58). Host cell transcriptome was recently inferred upon hMPXV infection (59). However, the transcriptome of hMPXV itself has not been analyzed.

Our objective in this study was to identify TSSs and TESs of viral mRNAs, which helps annotate the complete viral transcriptome. Furthermore, we identified the promoter and poly(A) signal consensus elements of the hMPXV genes.

Genome sequencing studies are essential for tracking genetic mutations during a viral outbreak, while transcriptomic analyses are necessary to discover novel genes. These analyses offer insights into gene expression patterns and isoform variations in poxviruses. Notably, non-conserved regions such as the inverted terminal repetition (ITR) segments undergo rapid microevolution (8, 60). These regions are crucial for determining host spectrum and evading host immune responses, making their study critically important. Despite a recent decline in clade IIb mpox cases, the risk of a future outbreak should not be underestimated. In 2023, the number of suspected hMPXV infections increased in the DRC, with genetic analysis indicating that clade I viruses, transmitted through sexual contact—a characteristic previously only described for clade IIb—are responsible (61).

## RESULTS

We employed two distinct sequencing approaches to identify the terminal regions of the hMPXV transcripts. TSSs were detected using CAGE-Seq on the Illumina MiSeq platform, whereas TESs were identified through dRNA-Seq on the ONT MinION device.

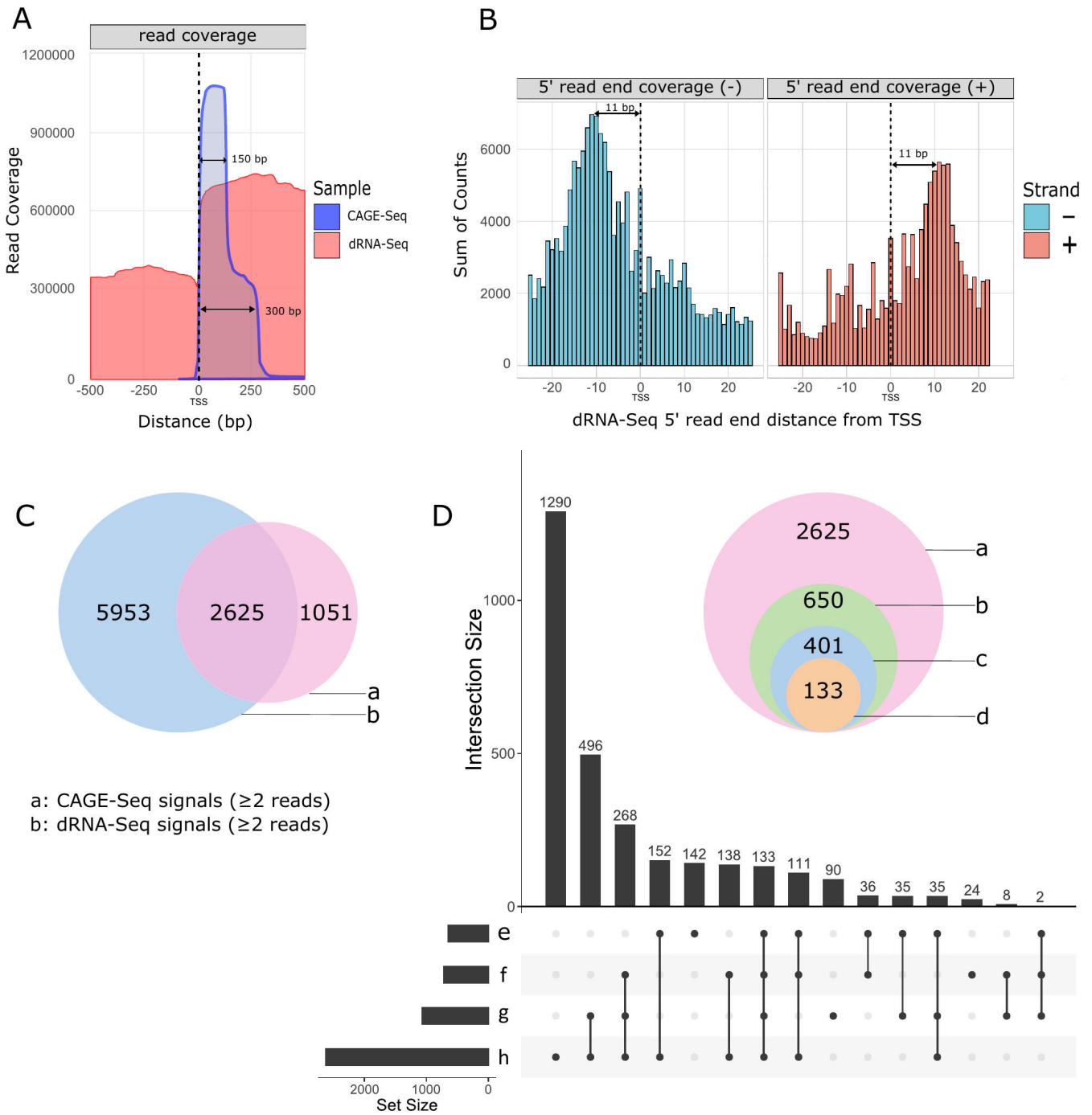
### Transcription start sites

CAGE-Seq analysis identified a total of 3,676 TSS positions excluding the singletons (Table S1), although dRNA-Seq efficiently validates 5'-ends but encounters challenges due to incomplete sequencing of these termini (53). However, unlike cDNA-Seq techniques, dRNA-Seq is free from common artifacts. Therefore, we opted to utilize this method to validate the results obtained from CAGE-Seq (Fig. 1A; Fig. S1). A total of 2,625 TSSs were confirmed by dRNA-Seq within a 25-nucleotide window, likely representing an underestimate of the overall TSSs. We analyzed the distribution of dRNA-Seq read ends in the proximity of CAGE-Seq signals and found that the 5'-ends detected by dRNA-Seq were most frequently positioned on an average of 11 nucleotides downstream from the TSSs identified by CAGE-Seq (Fig. 1B). The missing nucleotides at the 5'-end result from the premature release of the RNA molecules by the motor protein. We further filtered the 2,625 positions by eliminating those with fewer than 10 supporting reads, resulting in a total of 650 positions by excluding those supported by fewer than 10 reads (Fig. S2). Subsequently, we analyzed which of these positions were within 40 nt upstream of a predicted promoter. This latter analysis yielded a final count of 401 positions (Fig. 2).

Furthermore, employing another novel TSS clustering algorithm within the TSSr package (peakclu), we identified 646 clusters of CAGE signals, each with a single dominant peak (Table S2). Comparing these dominant peaks of the clusters with the data set of 401 filtered TSS positions, we identified a set of CAGE signals comprising 133 positions that met all the filtering criteria (Fig. 1C and D; Table S2).

This shows that both clustering and unclustering of CAGE signals lead to robust TSS detection, demonstrating their consistency. Using the shape score index, peak analysis of CAGE-Seq data revealed two major types of TSS distributions: broad and narrow range. The analysis indicated that the majority of the clusters consist of single peaks, with the vast majority the clusters not surpassing 10 nt in width (Fig. 3).

The distinguishing characteristic of poxviral mRNAs is the presence of poly(A) leader sequences at the 5'-ends of late mRNAs (62, 63). Despite the absence of 11 nt on average at the 5'-end of dRNA-Seq reads, the presence of a 5'-poly(A) leader enables the sequencing of the entire molecule, as shown in Fig. 3C. We estimated the number



**FIG 1** Distribution and characterization of 5'-ends of hMPXV mRNAs. (A) The figure shows the raw read coverage of all superimposed CAGE-Seq and dRNA-Seq reads around all annotated TSSs (dashed black line represents the position of TSSs). The x-axis represents the distance from the TSS, while the y-axis indicates the coverage. The CAGE-Seq is a composite of 150-bp and 300-bp libraries. The figure demonstrates that the coverage of dRNA-Seq and CAGE-Seq reads generally agrees, providing a clear signal for detecting transcriptional start positions. (B) The histogram illustrates the distribution of 5'-ends of dRNA-Seq reads around all CAGE-Seq TSSs in a  $\pm 25$ -nt window. The x-axis represents the distance from the TSS, while the y-axis indicates the sum of read counts. The dRNA-Seq 5'-ends most frequently accumulate 11 nt downstream from the TSSs, which is seen as two dominant peaks on the histogram. (C) Venn diagram shows the initial number of putative TSSs in CAGE-Seq and dRNA-Seq and their intersection before applying the filtering criteria. (D) The upset plot illustrates the intersections between different subsets of our data during the filtering process. The subsets are defined based on specific criteria applied to the CAGE-Seq and dRNA signals. The onion diagram depicts the number of CAGE-Seq signals according to the filtering method implemented in this study. (a) All detected CAGE-Seq peak signals, except singletons, corroborated by dRNA 5'-ends located within a 25-nucleotide window downstream from the TSS. (b) Number of CAGE-Seq peaks with at

(Continued on next page)

Fig 1 (Continued)

least 10 read counts corroborated by dRNA 5'-ends located within a 25 nucleotide window downstream from the TSS. (c) CAGE-Seq signals with at least 10 read counts, corroborated by a promoter motif detected within a 40-nucleotide interval and co-terminating with dRNA-Seq reads within a 25-nucleotide window downstream from the TSS. (d) Number of dominant TSS signals within the clusters of CAGE-Seq signals that match the filtered TSS data. (e) Clusters of CAGE-Seq signals that are the dominant peaks within their cluster. This subset focuses on the most prominent TSS within each cluster, indicating the primary sites of transcription initiation. (f) Set of CAGE-Seq signals that have a count value of at least 10. This subset is used to highlight robust CAGE-Seq signals that are likely to represent significant transcription start sites. (g) CAGE-Seq signals that have a promoter element within 40 nucleotides upstream. This subset helps identify TSSs that are located near promoter elements, providing insights into promoter activity. (h) CAGE-Seq signals that are validated by a dRNA 5'-end within 25 nucleotides. This subset indicates the proximity between CAGE-Seq and dRNA-Seq data, suggesting high-confidence TSSs.

of 5'-poly(A) leaders and found that 10% of CAGE-Seq reads and 5% of dRNA-Seq reads contain at least three A bases (Fig. 4).

TSS positions were sorted according to their abundance. The top five TSSs surpass a read depth of 1,000 in CAGE-Seq (Table S3). Among these, three TSSs stand out with exceptionally high CAGE-Seq signals, each showing count values exceeding 10,000. The highest CAGE-Seq signal represents 13% of the total and nearly 44% of the count for the top five TSSs. In dRNA-Seq, the gene OPG110, which encodes the protein VLTf-4 involved in post-replicative transcription elongation of L genes, has the most abundant 5'-end position. Out of the most abundant 5'-CAGE-Seq positions, three coincided with the most abundant dRNA-Seq positions belonging to the following genes: OPG065, OPGOPG110, and OPG022. Table S3 provides a summary of the orthologs and functions of genes associated with the most abundant TSS positions.

## Promoter elements

Our understanding of promoter elements in orthopoxviruses primarily stems from research on VACV (64, 65). Poxviruses use distinct promoter motifs in the early and late phases of infection (66). Given the close phylogenetic relationship between VACV and hMPXV (67, 68), the promoter motifs of the former virus were employed to identify corresponding elements in hMPXV (Table 1).

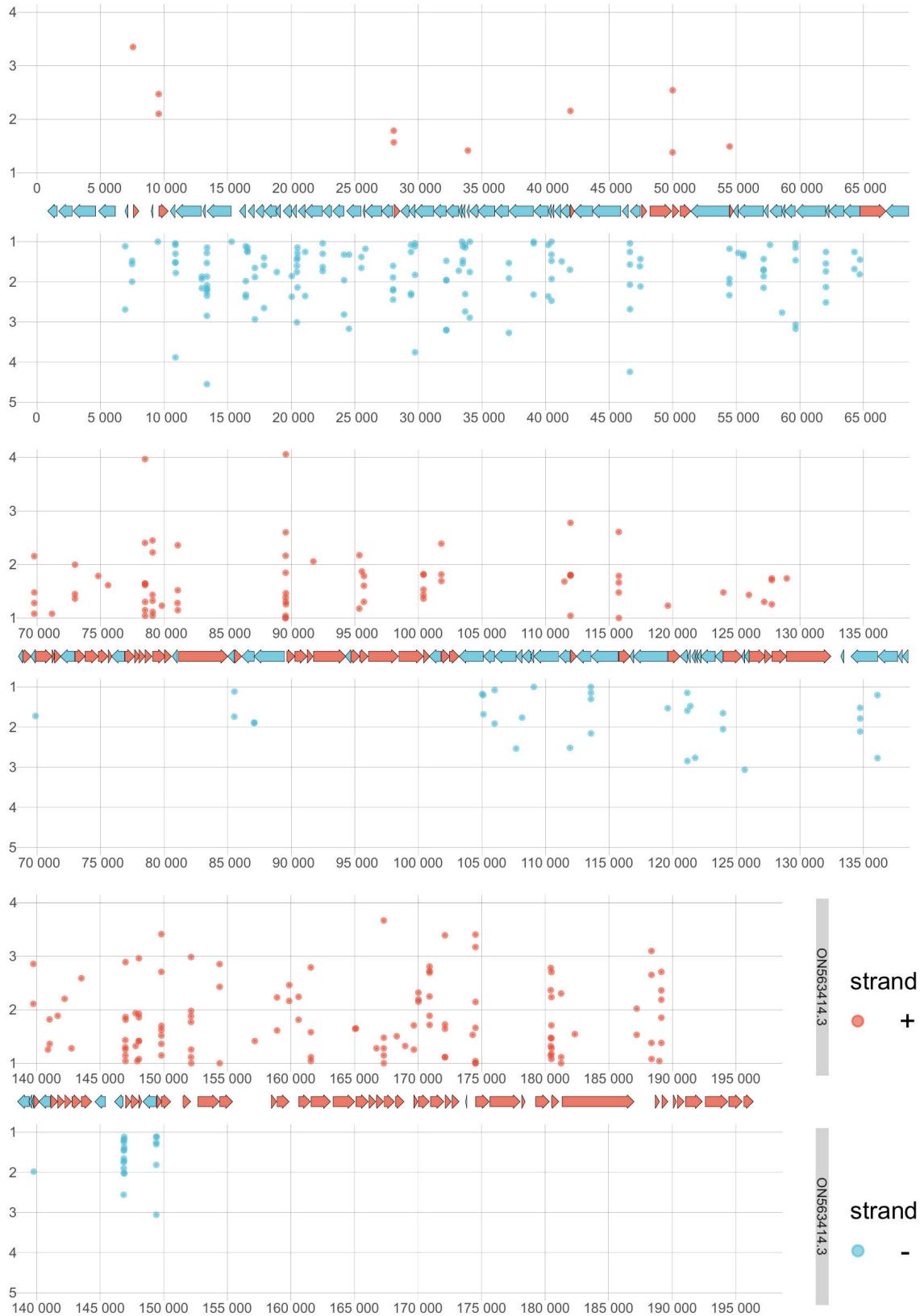
We identified 1,369 putative promoters within a 100-nt interval upstream of TSSs using the FIMO (Find Individual Motif Occurrences) program. The resulting predicted promoters, along with their *P* and *q*-values, are listed in Table S1c. The best-matching motifs, associated with the names of ORFs, are organized according to their *q*-values and detailed in Table S1d. The average distance between each TSS and its predicted promoter motif was determined to be 26 nucleotides, with the most frequent distance observed being 1 nucleotide (Fig. 5A). This finding is consistent with results from studies conducted on VACV (56).

## Transcription end sites

Direct RNA sequencing, based on poly(A) selection, was employed to identify the 3'-ends of hMPXV RNAs, using the LoRTIA (71) tool for TES annotation. A total of 3,241 positions were identified (excluding singlets), with 496 of these positions validated by a minimum of six reads (Fig. S3). Among these, 135 positions were further validated by ePAS signals within a 50-nucleotide distance (Fig. 6; Table S4).

## Poly(A) signals

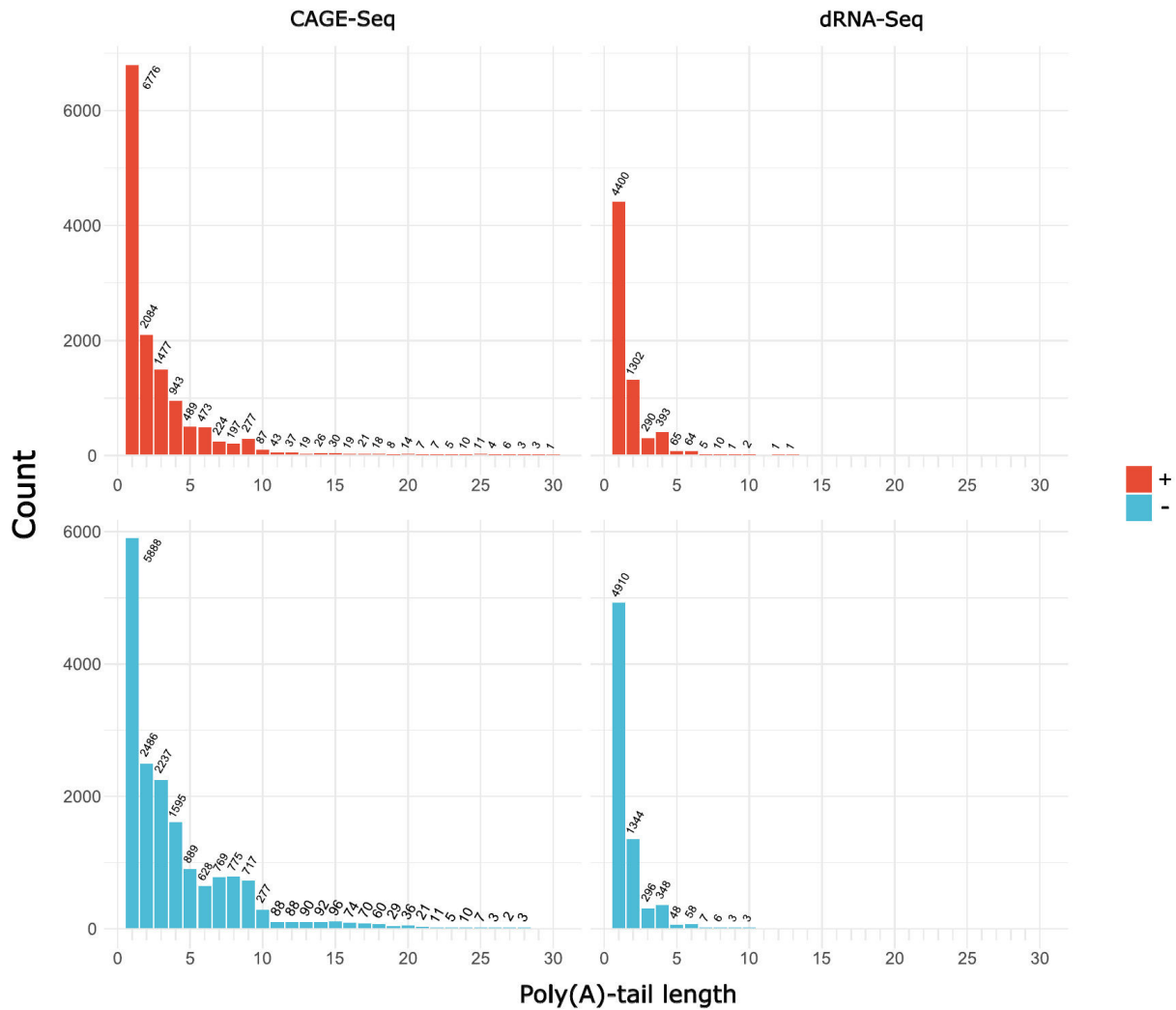
Orthopoxviruses utilize their unique enzymatic machinery to recognize polyadenylation signals (PASs) and to synthesize the poly(A)-tail of viral mRNAs. VACV early mRNAs are characterized by a UUUUUNU early PAS (ePAS), leading to a premature and homogenous end of early mRNAs (56, 72). Using a motif scanning algorithm (FIMO), we identified 734 ePASs, as detailed in Table S4. Of these, 313 ePASs were found 50 nt upstream of TESs, validating 135 of the previously mentioned 496 TESs, as reported in Table S4. The average distance of ePAS from TESs is 24 nt, which is in concordance with VACV data (56, 70). One benefit of dRNA-Seq is its ability to directly analyze the native poly(A)-tails of



**FIG 2** Distribution of filtered TSSs along the viral genome. The figure illustrates the annotated genome of hMPXV (ON563414.3), depicting the positions of TSSs determined by CAGE-Seq. We applied specific filtering criteria to identify these TSSs: a minimum of 10 CAGE-Seq signals at a position, a predicted promoter motif within a 40-nucleotide window upstream of the TSS, and at least one dRNA-seq 5'-end with a minimum read count of 2 within a 25-nucleotide window downstream from the TSS. This resulted in a total of 401 TSSs. TSSs on the positive strand are illustrated in red and those on the negative strand in blue. The x-axis denotes the values of CAGE-Seq peaks at each genomic position on a logarithmic scale, and the y-axis denotes the genomic positions.







**FIG 4** Distribution of the length of the 5'-poly(A) leader sequence in hMPXV. The distribution of the length of the 5'-poly(A) leader sequence in CAGE-Seq and dRNA-Seq samples from both the + and - strands. The x-axis denotes the length of the poly(A) leader (excluding values of 0), while the y-axis represents the number of reads.

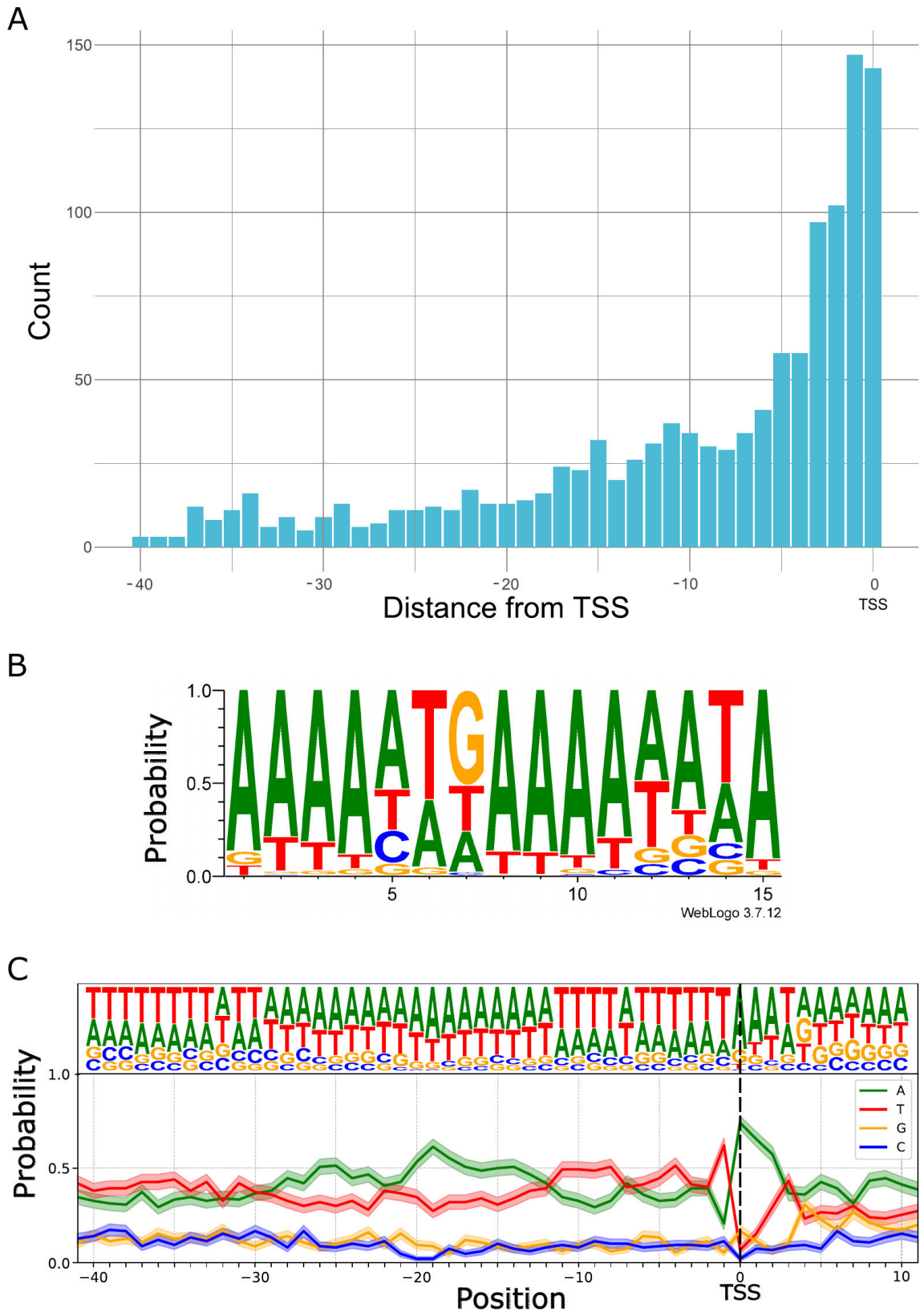
RNAs. In the analysis of 232,258 hMPXV mRNAs, the mean poly(A)-tail length was found to be 97.91 nt (with an SD of 51.07 nt) according to Nanopolish and 82.21 nt (with an SD of 43.48 nt) as measured by Dorado. The most frequent poly(A)-tail lengths were 86 nt and 71 nt (Fig. 7; Table S5).

**TABLE 1** Cis-regulatory sequences used for promoter and PAS annotation<sup>a</sup>

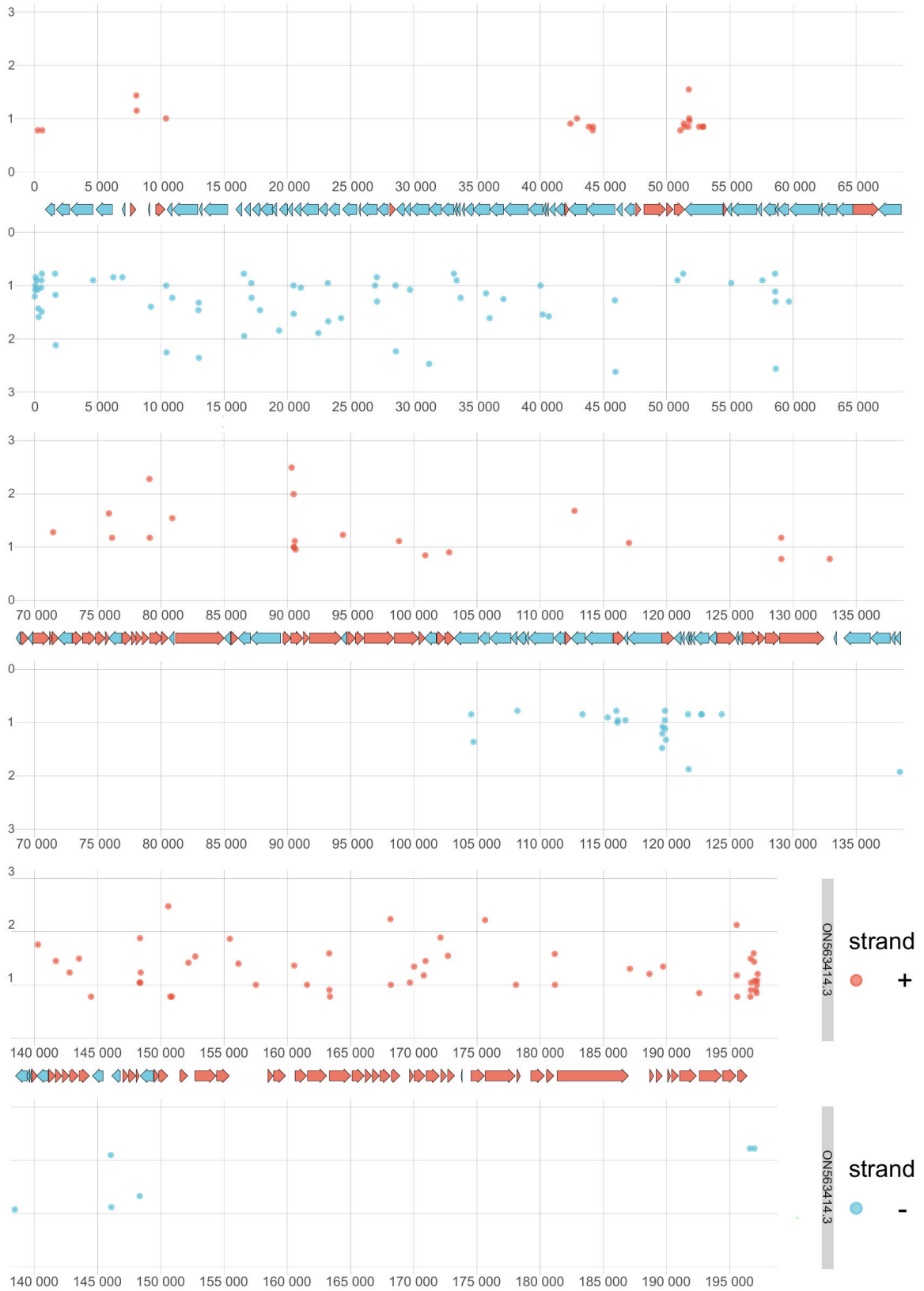
Type	Kinetics	Consensus	Reference
Promoter	Early	AAAANTGAAAANNA	56
	Late	TAAATG/NNNTNNNNNNNNNTAAATG	69; Yang et al. (56)
	Group I	NNNNNNNNNNYWNWWWWTWWNNNNNWTAAATG	Yang et al. (56)
	Group IIB	NATWWNNNNHTAAAAANNDNNNNHNDWWNTAAAYN	
	Group IIA	NRNNWNWTNWMWNWWWWTNDDNNNNH	
	Intermediate	NNNATNNNNNNNTAAAAANNNNNNNNNNNNTAA	70
	Mixed	NNNA/TNNNNNNNNNTNNNNNNNNNTAAATGGNNN	Yang et al. (56)
	Mixed	NTAWAD	Tombácz and Prazsák et al. (57)
PAS	Early	UUUUUNU	Yang, Reynolds et al. (70); Yang et al. (16)

<sup>a</sup>The promoter motifs used to scan viral promoter and PAS sequences are categorized by their kinetics, based on data from literature on experiments related to VACV gene expression. PAS stands for poly(A) signal.





**FIG 5** Promoter elements in hMPXV genome. (A) Distribution of promoter motifs within a 40-nt interval upstream of TSSs. (B) The consensus motifs of early promoters are illustrated by WebLogo. (C) Base composition probability near TSSs associated with post-replicative promoters. The TSS within the conserved TAAAT motif is indicated by dashed line.

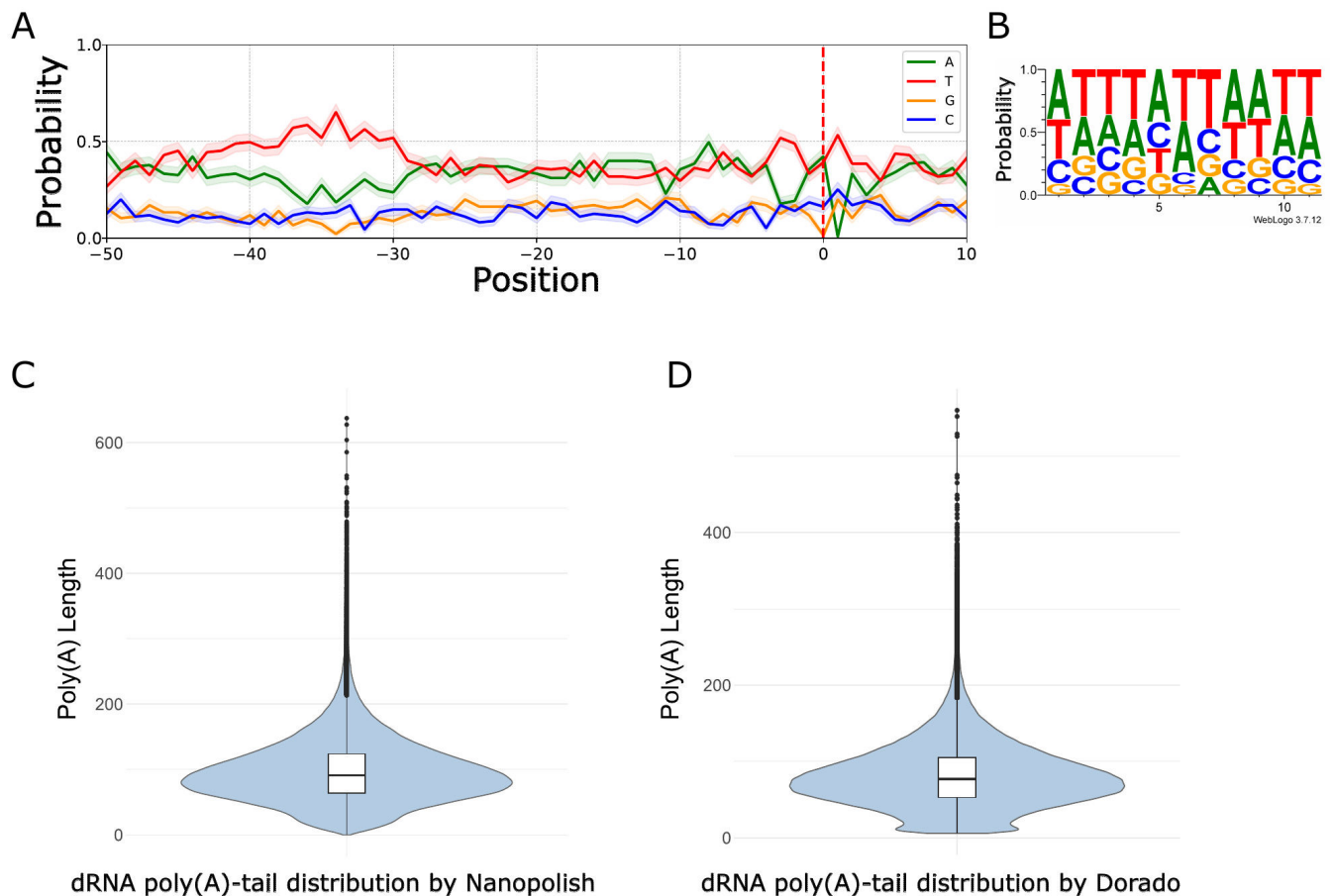


**FIG 6** Distribution of filtered TESs. The figure displays the global distribution of TES positions with a minimum count of six in the dRNA-Seq data. The x-axis represents the count on a log10 scale, while the y-axis indicates the genomic position.

### UTRs of hMPXV genes

The hMPXV genome displays the densely packed and sequentially arranged gene structure common to orthopoxviruses. This layout creates many short intergenic regions, with an average distance of 129 nucleotides between genes, which often causes the untranslated regions (UTRs) of neighboring genes to overlap. Following the annotation of TSSs and TESs, we identified the canonical UTR for each ORF in hMPXV. To determine the 5'-UTRs, we initially aligned the filtered TSS positions with the coordinates of a given ORF and selected the most abundant closest TSS as canonical.

We found that 118 out of 190 ORFs had an associated TSS, while the remainder either failed to meet our strict criteria or shared a common TSS with other ORFs. The length of the canonical 5'-UTRs ranged from 0 to 590 nt, with an average of 72 nt (see Table S6a). This excludes cases where the TSS was located within the host ORF. The 5'-UTRs can also be distinguished by their TSS distribution. We discovered that 63 ORFs have a single, highly abundant TSS, while 55 ORFs could be associated with a non-single peak type of TSSs. Additionally, 20 TSSs were found in the upstream neighboring ORF, and 7 TSSs were detected within the host ORF (Table S6b). The dRNA-Seq facilitates the identification of complete transcript boundaries. We observed that many ORFs have alternative transcription initiation sites, often including one or more upstream ORFs. We identified these low-abundance 5'-UTRs, where one or more ORFs are covered by the UTR of a downstream ORF, as detailed in Table S6a. It is known that VACV produces heterogeneous 3'-ends (16, 27); therefore, determining the length of 3'-UTRs is challenging. We examined the 3'-UTRs of hMPXV based on the closest TES to a given ORF and found that



**FIG 7** Characterization of TESs and poly(A)-tails of hMPXV mRNAs. (A) The PASs of the early ORFs are located within 50 nt upstream of the TESs, which are represented by a dashed red line. (B) The TES is characterized by a dominant A/T nucleotide composition. (C) The poly(A)-tail length distribution of viral dRNA-Seq reads estimated by Nanopolish. (D) The poly(A)-tail length distribution of viral reads estimated by Dorado.

out of 190 ORFs, 113 are assigned to TEs. The mean length of 3'-UTRs was found to be 176 nt. According to our data, almost half of the canonical 3'-UTRs are terminated in the downstream ORFs (Table S7).

### Putative novel genes

An in-depth analysis of TSS positions showed CAGE-Seq signals within intergenic spaces located at the variable ends of the genome. These signals, identified in both the right and left terminal regions, were validated by the ends of dRNA-Seq reads (Table 2).

The new genes were further corroborated by the prediction of promoter elements and by dRNA-Seq identifying their TEs (Table S8). Three of the most abundant novel TSSs are demonstrated on Fig. 8.

Despite their short predicted ORFs, a pBLAST search revealed homology with poxviral sequences for three entities: a hypothetical ankyrin-repeat containing protein (located between OPG210 and OPG214), a Kelch-like motif containing a possible protein-coding sequence (located between OPG181 and OPG184), and another unknown protein-coding gene situated in the intergenic area of OPG015 and OPG018 (Fig. 8; Table S8).

Nucleotide BLAST was employed to identify orthologs of the novel genes among orthopoxviruses. Five out of six novel genes matched known orthologs of other orthopoxviruses. The exception is found in the intergenic region between OPG020 and OPG021. Detailed results from the nBLAST analysis are provided in Table S8.

### Phylogenomic analysis identifies the MPXV isolate in the B.1 lineage of clade IIb

We performed phylogenomic analysis for the determination of the evolutionary relationship of our isolate. Based on the distance matrix and SNP analysis of the alignment of 44 representative members of MPXVs, our virus isolate belongs to the B.1 lineage of clade IIb (Supplementary File 1).

## DISCUSSION

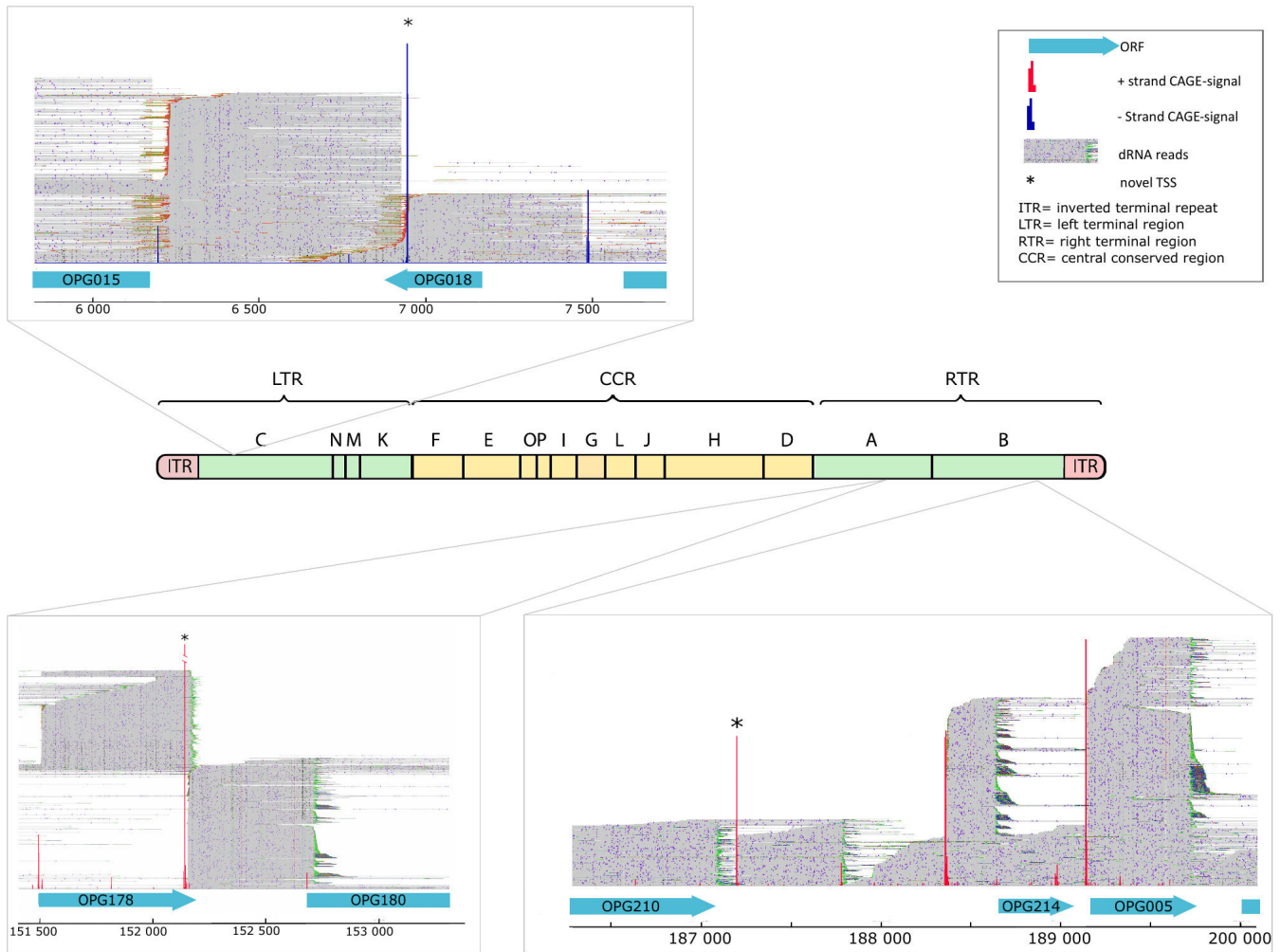
Although the human monkeypox virus has been known for decades and has sporadically caused periodic outbreaks in Eastern, Central, and Western Africa, the 2022 outbreak has heightened awareness of the virus. Recent studies suggest that the virus is undergoing rapid microevolutionary changes. Within just a few years, both clade I and clade II have evolved the ability to spread from human to human. Therefore, it is crucial to investigate and understand these clades more thoroughly at both the genomic and transcriptomic levels. While many studies have examined hMPXV at the genomic level, they have often missed detecting novel transcriptionally active sites that represent potential new genes. In this study, we employed dRNA-Seq on the ONT MinION platform to identify the precise TEs of hMPXV, known for their considerable diversity in poxviruses (57). Detection of poly(A) signals was used for the validation of dRNA-Seq results. The lRNA-Seq methods, particularly of the ONT approach, have been found to produce a pervasive

**TABLE 2** List of novel TSS and TES positions in intergenic region of hMPXV<sup>a,b</sup>

TSS (CAGE)	TES (dRNA)	Strand	Localization	Promoter start	Adjacent downstream gene	Adjacent upstream gene
6,936	6,230	–	LTR	6,930; 6,949	OPG018	OPG015
9,501	9,203	–	LTR	9,504	OPG020	OPG021
152,144	152,157	+	RTR	152,117	OPG178	OPG180
157,160	157,506	+	RTR	157,126	OPG181	OPG184
168,981	169,692	+	RTR	168,951	OPG195	OPG197
187,189	187,794	+	RTR	187,160	OPG210	OPG214

<sup>a</sup>Novel TSSs and TESs have been identified in both the left and right variable regions of the hMPXV genome. Their positions were determined based on sequence alignment against the first public hMPXV reference sequence (ON563414.3) from the 2022 outbreak (73). The locations of the TSSs are indicated as follows: left terminal region (LTR) and right terminal region (RTR). The possible lengths of ORFs are calculated by taking the coordinates from the first ATG to the following STOP codon, along with the dRNA-seq reads.

<sup>b</sup>LTR= left terminal region, RTR= right terminal region.



**FIG 8** Novel hMPXV genes. The figure shows the localization of the three most abundant novel genes in the hMPXV genome. These putative novel genes are located within intergenic positions. ORFs are indicated with blue boxes in both the right and left terminal regions of the genome. Novel TSSs are indicated by asterisks. The dRNA reads visualized in IGV reveal a novel gene located between OPG18 (MA001-005/D2L) and OPG015 (MA001-004) at the left terminal region. A novel TSS is located between the ORFs OPG178 (MA001-158/Thymidylate kinase) and OPG180 (MA001-159/DNA-ligase) and downstream of the OPG210 (MA001-182/B21R) gene in the right terminal region of hMPXV. The letters above the genome indicate the HindIII fragments of hMPXV (source: ViralZone).

5'-truncation of transcripts, potentially leading to incorrect identification of false TSSs (74). Our previous investigations (18, 75, 76) have also uncovered a diverse range of 5'- and 3'-transcript ends in various viruses, many of which, particularly the TSSs, are likely non-functional or could even be of non-biological origin. To address this issue, we employed CAGE sequencing on an Illumina MiSeq platform, a well-established method for detecting the 5'-ends of capped RNA molecules. While CAGE-Seq is highly reliable, we cannot exclude the possibility that a certain fraction of degraded RNA molecules is also detected by this technique, since it has been shown that mammalian cells contain enzymes in the cytoplasm capable of generating Caps onto uncapped RNAs (77). A key issue is the absence of software capable of unequivocally differentiating genuine RNA molecules from technical artifacts. In light of this, our study focused on the annotation of main transcript ends but also provided data on the low-abundance putative TSSs and TESs.

We compared the 5'-ends of mRNAs from CAGE-Seq libraries, to those generated by dRNA-Seq, and detected that a significant portion of dRNA-read ends is accumulated on an average of 11 nt downstream of a TSSs (Fig. 1B). This discrepancy is mainly due to



poor-quality ends of dRNA-Seq reads, which fail to align when local alignment methods are used. To overcome this phenomenon, SRS and LRS methods are combined (78, 79), or adapter ligation is carried out (80).

VACV is the best-studied representative of orthopoxviruses. Since VACV and hMPXV are phylogenetically closely related (81), their promoter motifs are expected to be very similar. Therefore, we scanned the hMPXV genome using a set of VACV promoter modules. The validation of TSSs and TESs was carried out by identifying nearby consensus sequences and poly(A) signals, respectively. We also compiled a list of high-abundance putative transcript ends where cis-regulatory sequences could not be identified nearby. Integration of short- and long-read sequencing data provided a high-resolution map of the viral transcript ends. Extremely high levels of transcriptional activity were detected in both the core and terminal regions of the viral genome. Additionally, we observed mRNA readthrough at the peak of the circularized genome. The positions of the most abundant TSSs, along with their corresponding host ORF, and their VACV orthologs are listed in the Table S3. The most abundant TSS belongs to the gene OPG023, which codes for a short, non-essential protein termed D8L containing an ankyrin-like peptide domain. This domain plays a role in host immune evasion by blocking IL-1 receptors (82) and modulating the NF- $\kappa$ B pathway (83). The second most abundant TSS belongs to the gene OPG065, which might have evolved via episodic positive selection in response to immune selection (68) and host antiviral response (84). In the dRNA-Seq analysis, the most abundant TSS is associated with the hMPXV OPG110 gene, which plays a critical role in replication and for virion morphogenesis (85, 86).

Our findings on the TSS pattern align with previous studies, confirming the existence of two major TSS types: single-peak and broad-range CAGE-Seq signal distributions. Similar patterns have been observed in orthopoxviruses (56), Herpesviruses (79), and other organisms (87, 88). More precise mapping of the TSSs and additional mutagenesis studies are needed to further explore the transcriptomic structure of poxviruses.

Termination of poxvirus transcription requires the interaction between a U(5)NU consensus sequence and the assembly of a ternary complex, which includes the viral termination factor and the RAP94 protein, causing strict 3'-termination of transcripts (89, 90). Unlike early mRNAs, PR RNAs exhibit high heterogeneity in length because the ePAS is unrecognized by the poxvirus transcription termination complex (16, 56). The transcription of orthopoxvirus genes often terminates within the downstream ORFs (16, 56).

Using oligodT selection-based library screening, canonical TES positions were assigned to the annotated ORFs. However, our dRNA-Seq analysis shows that not all ORFs can be assigned canonical TESs due to the presence of TESs likely used by more than one gene in hMPXV. A similar pattern of TES distribution was revealed in VACV using LRS (27, 57), suggesting the formation of co-terminal transcription units. Our LRS method also enabled the annotation of 73 ePAS, confirming the existence of early canonical TESs (Table S4). We detected a 3'-UTR architecture similar to VACV in the hMPXV transcriptome.

We found that the average length of 5'-UTRs in hMPXV is short, consistent with findings reported by others for other orthopoxviruses (56, 57). In some rare cases (Table S6b), anomalous TSSs were located downstream to the annotated start codon, suggesting alternative ATG usage by the virus (27, 84).

It is important to note that UTRs can be influenced by insertion/deletion events. For example, downstream of the OPG201 (MA001-175) gene, only a diffuse PAS is detectable, and the OPG202 (MA001-176) gene lacks a TES. This region contains an [ATAT] repeat, which can disrupt mRNA termination signals (8). However, a detailed analysis of low-complexity regions in our transcriptome revealed that some TESs could not be precisely determined. This could be due to the presence of repeats or the so-called "chaotic" transcriptomic regions, independent of the kinetic class of transcripts, as described in the transcriptomic analysis of VACV-WR (27). The UTR length in poxviruses is influenced by the kinetic expression of a given transcript. It is known that early



mRNAs have homogeneous 3'-ends, while post-replicative mRNAs exhibit heterogeneous UTR lengths (16, 56). The presence of 5'-poly(A) leader is a characteristic feature of the poxviral mRNAs (16). Furthermore, VACV is a cytoplasmic virus, possessing two enzymes (D9, D10) functioning as decapping enzymes in mRNA degradation and translation regulation. In our study, we also detected the poly(A) leaders in both the dRNA and CAGE samples. Although literature suggests an average length of 35 nt for these sequences (91), we observed shorter lengths in hMPXV. However, it is important to consider that these shorter lengths may be underestimations due to the possible incomplete sequencing of the 5'-end.

Direct RNA sequencing confirmed the presence of polyadenylated novel mRNAs in intergenic regions of both the left and right ITRs. These regions of poxviruses are thought to be responsible for host-virus interactions (8); therefore, a similar function is expected for the novel genes. Farlow and colleagues (92) reported mutations in a cidofovir-resistant MPXV strain in the same genomic region. They speculated about the presence of a hypothetical yet-unknown ankyrin-like protein-coding gene which we can confirm here. On the other hand, this virus is classified within the clade II B.1 lineage of hMPXV. Phylogenetic studies show a relatively high mutation rate within this lineage (93, 94). This accelerated evolution is suggested to be driven by the action of the cellular APOBEC3 nucleic acid-editing enzyme in the terminal genomic region (95–97). Genotyping hMPXV via gene or genome sequencing and identifying point mutations are frequently employed to track the pandemic's progression (12, 98). Several studies have aimed to elucidate the pathogenicity and virulence of hMPXV by examining variations in the terminal region, which encodes proteins involved in immune modulation (99–101). Nonetheless, transcriptomic studies provide the benefit of describing the functional units of the viral genome, rather than merely analyzing gene variants.

The presence of tandem repeats in the ITR regions of poxvirus genomes is well documented (102–104). Desingu et al. identified a region of tandem repeats (AACT AACTTATGACTT) in the 5'-ITR and 3'-ITR regions of the hMPXV (clade IIb B.1) virus, which is absent in other poxviruses (105). Although the function of these unique repeat sequences remains unknown, gene loss and gain have been observed at the ends of the 5'-ITR and 3'-ITR regions among clade I, clade IIa, and clade IIb mpox viruses (106), indicating a continuous mutational hot spot for the virus (105). These changes are accompanied by unique tandem repeats (8).

These regions are located between the ORFs MPXVgp003 and MPXVgp004, positioning them as intergenic and potentially good targets for foreign proteins in vaccine development. However, our long-read dRNA-sequencing data reveal strong transcriptional activity in these tandem repeat regions, suggesting they are not truly intergenic.

In conclusion, we employed advanced sequencing techniques to comprehensively map the transcript ends and cis-regulatory elements of hMPXV. By combining long- and short-read sequencing methods, we accurately identified TSSs, TESs, and their promoter elements. Additionally, we discovered six potential new genes, significantly updating the genic and intergenic annotation of hMPXV. Our findings underscore the importance of ongoing transcriptomic exploration in infectious disease research, emphasizing the need for further studies to elucidate the dynamic transcription profile of the virus and its complex interactions with the host.

## MATERIALS AND METHODS

### Virus propagation and RNA isolation

The methods for cell culture, virus propagation, and RNA isolation are detailed in the Supplemental Text. Briefly, the hMPXV isolate was propagated in CV-1 cell lines at a multiplicity of infection (MOI) of 5, with three replicates, in 75-cm<sup>2</sup> flasks. The infected cells were then incubated at 37°C for 2, 6, 12, and 24 hours. RNA was isolated using the Nucleospin RNA Mini Kit (Macherey Nagel) according to the manufacturer's protocol at each time point, followed by DNase treatment to remove residual DNA. Thereafter,

polyadenylated RNA enrichment was carried out using Lexogen's Poly(A) RNA Selection Kit V1.5. RNA samples were bound to beads, washed, and hybridized. After incubation and washing, the polyadenylated RNA was eluted in nuclease-free water and stored at  $-80^{\circ}\text{C}$  for subsequent analysis.

## Native RNA sequencing

The Oxford Nanopore Technologies SQK-RNA002 Kit was utilized to sequence the RNA molecules. For library preparation, we used 50 ng (in 9  $\mu\text{L}$ ) of a pooled sample of poly(A)<sup>(+)</sup> RNAs. The initial step involved the ligation of a 1- $\mu\text{L}$  RT Adapter (110 nM; part of the ONT Kit) to the RNA sample using a mix of 3  $\mu\text{L}$  NEBNext Quick Ligation Reaction Buffer (New England BioLabs), 0.5  $\mu\text{L}$  RNA CS (ONT Kit), and 1.5  $\mu\text{L}$  T4 DNA Ligase (2 M U/mL, New England BioLabs). This process was conducted at room temperature (RT) for 10 minutes. Subsequently, the cDNA strand was synthesized using SuperScript III Reverse Transcriptase (Life Technologies), with the reaction taking place at  $50^{\circ}\text{C}$  for 50 minutes, followed by a 10-minute inactivation phase at  $70^{\circ}\text{C}$ . After this, the sequencing adapters from ONT's DRS Kit were ligated to the cDNA at RT for 10 minutes using the T4 DNA ligase enzyme and NEBNext Quick Ligation Reaction Buffer. The final direct RNA library was sequenced on an R9.4 SpotON Flow Cell. To wash the direct RNA-seq and direct cDNA-seq libraries after each enzymatic reaction, RNAClean XP beads and AMPure XP beads (both sourced from Beckman Coulter) were employed.

## Cap analysis of gene expression

The detailed protocol is described in the Supplemental Methods. Briefly, to investigate TSS patterns in hMPXV, we used CAGE-Seq. Total RNA (5  $\mu\text{g}$ ) was prepared into CAGE-Seq libraries, starting with RNA denaturation and first-strand cDNA synthesis using the CAGE Preparation Kit. Post synthesis, the RNA was oxidized, and biotin was attached to the 5'-Cap. Biotinylated RNA underwent Cap-trapping on Streptavidin beads, followed by sequential washing and cDNA release. The capped cDNAs were isolated and treated with RNase to remove residual RNA. Streptavidin beads were prepared and washed, and linkers were attached to the cDNAs. After ligation, samples were treated with Shrimp Alkaline Phosphatase and USER enzyme to prepare for second-strand cDNA synthesis. Following synthesis, the samples underwent multiple purification steps and were sequenced on an Illumina MiSeq instrument. The sample concentration and library quality were assessed using Qubit 4.0 and TapeStation, ensuring accurate transcription start site profiling. The CAGE sequencing was performed on the MiSeq platform with the v2 (using 150 cycles) and v3 (using 300 cycles) reagent kit.

## Bioinformatics

### CAGE sequencing analysis

The reads derived from CAGE-Seq were mapped using STAR to the reference genome with the following parameters: `STAR --runThreadN 8 --outSAMunmapped Within --alignIntronMax 1000`. The bam files were merged after mapping into one dataset (Fig. S4). The downstream analysis was conducted within an R environment (version: 4.2). Due to technical artifacts and stochastic transcriptional activities, TSSs inferred from CAGE-Seq may not represent *bona fide* TSSs. Therefore, we applied the TSSr program (<https://github.com/Linlab-slu/TSSr>) for CAGE-Seq signal analysis, which effectively handles this problem (107). As one function of TSSr did not work properly, we removed the soft-clips from the alignments using the script at GitHub (<https://github.com/gabor-gulyas/softclipremover>). The `getTSSs` function was used with two sets of parameters: one for the core region and one for the repeat regions. In the core region, default parameters were used; however, the threshold for the mapping quality in the terminal repeats needed to be decreased ( $mapq \geq 3$ ) to include the secondary alignments that have lower values. The distribution of CAGE-signals has been calculated by the SI score of TSSr's `shapeCluster` function. TSS clusters were identified by the "`peakclu`" algorithm in TSSr.

The *clusterTSS* function calculates the inter-quantile width of TSS clusters based on the cumulative distribution of CAGE signals. At least 80% of CAGE signals within a cluster was defined as the 5'- and 3'-boundaries of the TSS clusters (107).

### Long-read direct RNA sequencing analysis

During sequencing, the reads generated were basecalled using the fast model of the Guppy program (<https://community.nanoporetech.com>). We performed the mapping using Minimap2 (version: 2.17-r941) with the following parameters: `minimap2 -ax splice -Y -C5 -t4 --cs`. The reference genome was downloaded from NCBI GenBank (accession: [ON563414.3](https://.ncbi.nlm.nih.gov/nuccore/ON563414.3)) (73). Furthermore, we used the LoRTIA pipeline, developed in our laboratory, for assessing sequencing adapter quality and poly(A) sequences. It also helps eliminate false TESs that could arise from several sources, as described earlier (47). To ensure the alignments were not results of internal priming events, we applied the `talon_label_reads` submodule of the TALON software package (108).

The LoRTIA program (<https://github.com/zsolt-balazs/LoRTIA>) was used with the following parameters: `LoRTIA five_score = 16.0, three_score = 16.0 three_adapter='AAAA AAAAAAAAAA', five_adapter='GCTGATATTGCTGGG'` to identify 5'- and 3'-adapters on the sequencing reads and to determine the TES positions. To estimate the length of polyA tails of viral native RNAs, two methods were used: (i) Nanopolish (<https://github.com/jts/nanopolish>) using the `polyA` command with default parameters and (ii) Dorado (<https://github.com/nanoporetech/dorado>) using the following parameters: `--estimate-poly-a --min-qscore 6`.

### Identifying the promoter elements and poly(A) signals of hMPXV

These sequence elements were identified using FIMO (109). For promoter identification, the following command was used: `fimo --oc . --verbosity 1 --bgfile --nrdb-- --thresh 1.0E-4 motifs.meme ON563414.3.fasta`, while for PAS identification, the same command was used with the exception of lowering the threshold to  $10^{-3}$  (`--thresh 1.0E-3`).

### Poly(A)-tail length estimation

We implemented poly(A)-tail length estimator packages from Nanopolish (110) and Dorado (v0.5.3) to retrieve the length of poly(A)-tails of viral mRNAs. While the 5'-poly(A) leader sequences were counted at the 5'-soft-clipped region of mapped mRNAs allowing one mismatch after three bases of As/Ts.

### Clade determination of the viral isolate

We performed genome alignment of our isolate (MPXV\_NRL\_4279\_2022) with the consensus sequence obtained from dRNA-Seq reads using the MAFFT algorithm (111) with the following settings: gap penalty open: 1.53, offset value: 0.123, and scoring matrix: 200 PAM/ $k = 2$ . Representative members of each MPXV clade were downloaded from the NCBI viral genome collection based on published data (105, 112). The classification of our sample was determined based on the distance matrix, which was calculated using Geneious software (Supplementary File 1).

### ACKNOWLEDGMENTS

We would like to thank Ferenc Jakab for his intellectual contributions to the preparation of this article and Fanni V. Földes, Brigitta Zana, and Zsófia Lanszki for their assistance in propagating the virus.

The research was funded by the National Research, Development and Innovation Office (NRDIO), through the researcher-initiated research projects (Grant number: K 142674) awarded to Z.B. G.Á.N. was supported by the New National Excellence Program (ÚNKP-23-3-SZTE-306) of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. G.E.T. and G.K. were supported

by the National Research, Development and Innovation Office, Hungary under grant RRF-2.3.1-21-2022-00010. The Article Processing Charges (APC) were paid by the Open Access Fund of the University of Szeged: 6683.

### AUTHOR AFFILIATIONS

<sup>1</sup>Department of Medical Biology, Albert Szent-Györgyi Medical School, University of Szeged, Szeged, Hungary

<sup>2</sup>National Laboratory of Virology, Szentágotthai Research Centre, University of Pécs, Pécs, Hungary

<sup>3</sup>Institute of Biology, Faculty of Sciences, University of Pécs, Pécs, Hungary

<sup>4</sup>Bernhard Nocht Institute for Tropical Medicine, WHO Collaborating Centre for Arbovirus and Hemorrhagic Fever Reference and Research, Hamburg, Germany

<sup>5</sup>Veterinary Research Institute, Brno, Czechia

<sup>6</sup>Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czechia

<sup>7</sup>Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia

### AUTHOR ORCID*s*

Dóra Tombác  <http://orcid.org/0000-0001-5520-2978>

Zsolt Boldogkői  <http://orcid.org/0000-0003-1184-7293>

### FUNDING

Funder	Grant(s)	Author(s)
<a href="#">Nemzeti Kutatási Fejlesztési és Innovációs Hivatal (NKFI)</a>	K 142674	Zsolt Boldogkői
<a href="#">Nemzeti Kutatási Fejlesztési és Innovációs Hivatal (NKFI)</a>	RRF-2.3.1-21-2022-00010	Gábor Kemenesi

### AUTHOR CONTRIBUTIONS

Gergely Ármin Nagy, Data curation, Formal analysis, Visualization, Writing – original draft | Dóra Tombác, Investigation, Methodology, Writing – review and editing | Zsolt Csabai, Investigation, Methodology | Ákos Dörmő, Investigation, Methodology | Gábor Gulyás, Data curation, Formal analysis | Gábor Kemenesi, Conceptualization, Investigation, Resources, Supervision, Writing – review and editing | Gábor E. Tóth, Investigation, Methodology, Writing – review and editing | Jiří Holoubek, Investigation, Methodology, Writing – review and editing | Daniel Růžek, Investigation, Methodology, Writing – review and editing | Balázs Kakuk, Data curation, Formal analysis, Writing – review and editing | Zsolt Boldogkői, Conceptualization, Formal analysis, Investigation, Supervision, Validation, Writing – original draft, Writing – review and editing.

### DIRECT CONTRIBUTION

G.Á.N. performed bioinformatics, analyzed and interpreted the data, and produced the figures. D.T. carried out CAGE-seq experiments, participated in experiment design and data analysis. I.P. engaged in data analysis and writing the manuscript. Z.C. isolated total RNA and took part in dRNA sequencing. Á.D. contributed to RNA isolation and executed poly(A)-selection and direct RNA sequencing. G.G. contributed to data analysis and bioinformatics. G.K. conducted viral infection. G.E.T. conducted viral infection. J.H. cultivated the virus. D.R. cultivated the virus. B.K. participated in bioinformatics. Z.B. conceived and designed the experiments and supervised the project. All authors reviewed and approved the final paper.

## DATA AVAILABILITY

Bam files from CAGE-Seq have been deposited in the European Nucleotide Archive and are available under the Project Accession: [PRJEB60061](https://www.ebi.ac.uk/ena/browser/view/PRJEB60061). dRNA-Seq data are available from the [PRJEB56841](https://www.ebi.ac.uk/ena/browser/view/PRJEB56841) study.

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**File\_S1 (mSphere00356-24-s0001.csv)**. Distance matrix of aligned MPXV genomes.

**Supplemental Figures (mSphere00356-24-s0002.pdf)**. Figures S1 to S4.

**Supplemental Material (mSphere00356-24-s0003.docx)**. Legends to supplemental material and supplemental methods.

**Table S1 (mSphere00356-24-s0004.xlsx)**. CAGE-Seq peaks detected by TSSr.

**Tables S2 and S3 (mSphere00356-24-s0005.xlsx)**. TSSs and shape scores.

**Table S4 (mSphere00356-24-s0006.xlsx)**. TESs.

**Table S5 (mSphere00356-24-s0007.csv)**. Estimated poly(A)-tail lengths.

**Table S6 (mSphere00356-24-s0008.xlsx)**. 5'-UTRs.

**Table S7 (mSphere00356-24-s0009.xlsx)**. 3'-UTRs.

**Table S8 (mSphere00356-24-s0010.xlsx)**. Novel TSSs found in intergenic regions,

## REFERENCES

1. Diven DG. 2001. An overview of poxviruses. *J Am Acad Dermatol* 44:1–16. <https://doi.org/10.1067/mjd.2001.109302>
2. Moss B. 2013. Poxvirus DNA replication. *Cold Spring Harb Perspect Biol* 5:a010199. <https://doi.org/10.1101/cshperspect.a010199>
3. Moss B, Smith GL. 2021. Poxviridae: the viruses and their replication, p 573–613. In Howley PM, Knipe DM, Cohen JL, Damania BA (ed), *Field's virology*, 7th ed. Wolters Kluwer.
4. Elwood JM. 1989. Smallpox and its eradication. *J Epidemiol Community Health* 43:92–92. <https://doi.org/10.1136/jech.43.1.92>
5. Ulaeto D, Agafonov A, Burchfield J, Carter L, Happi C, Jakob R, Krpelanova E, Kuppalli K, Lefkowitz EJ, Mauldin MR, de Oliveira T, Onoja B, Otieno J, Rambaut A, Subissi L, Yinka-Ogunleye A, Lewis RF. 2023. New nomenclature for mpox (monkeypox) and monkeypox virus clades. *Lancet Infect Dis* 23:273–275. [https://doi.org/10.1016/S1473-3099\(23\)00055-5](https://doi.org/10.1016/S1473-3099(23)00055-5)
6. Adler H, Gould S, Hine P, Snell LB, Wong W, Houlihan CF, Osborne JC, Rampling T, Beadsworth MB, Duncan CJ, et al. 2022. Clinical features and management of human monkeypox: a retrospective observational study in the UK. *Lancet Infect Dis* 22:1153–1162. [https://doi.org/10.1016/S1473-3099\(22\)00228-6](https://doi.org/10.1016/S1473-3099(22)00228-6)
7. Ladnyj ID, Ziegler P, Kima E. 1972. A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. *Bull World Health Organ* 46:593–597.
8. Monzón S, Varona S, Negredo A, Vidal-Freire S, Patiño-Galindo JA, Ferrusini-Gerpe N, Zaballos A, Orviz E, Ayerdi O, Muñoz-Gómez A, et al. 2024. Monkeypox virus genomic accord strategies. *Nat Commun* 15:3059. <https://doi.org/10.1038/s41467-024-46949-7>
9. Americo JL, Earl PL, Moss B. 2023. Virulence differences of mpox (monkeypox) virus clades I, IIa, and IIb.1 in a small animal model. *Proc Natl Acad Sci U S A* 120:e2220415120. <https://doi.org/10.1073/pnas.2220415120>
10. Rezza G. 2019. Emergence of human monkeypox in west Africa. *Lancet Infect Dis* 19:797–799. [https://doi.org/10.1016/S1473-3099\(19\)30281-6](https://doi.org/10.1016/S1473-3099(19)30281-6)
11. Damon IK. 2011. Status of human monkeypox: clinical disease, epidemiology and research. *Vaccine (Auckl)* 29 Suppl 4:D54–D59. <https://doi.org/10.1016/j.vaccine.2011.04.014>
12. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, Mixão V, Ferreira R, Santos D, Duarte S, Vieira L, Borrego MJ, Nuncio S, de Carvalho IL, Pelerito A, Cordeiro R, Gomes JP. 2022. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med* 28:1569–1572. <https://doi.org/10.1038/s41591-022-01907-y>
13. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. 2010. Orthopoxvirus genome evolution: the role of gene loss. *Viruses* 2:1933–1967. <https://doi.org/10.3390/v2091933>
14. Walsh D. 2017. Poxviruses: slipping and sliding through transcription and translation. *PLoS Pathog* 13:e1006634. <https://doi.org/10.1371/journal.ppat.1006634>
15. Baldick CJ, Keck JG, Moss B. 1992. Mutational analysis of the core, spacer, and initiator regions of vaccinia virus intermediate-class promoters. *J Virol* 66:4710–4719. <https://doi.org/10.1128/JVI.66.8.4710-4719.1992>
16. Yang Z, Martens CA, Bruno DP, Porcella SF, Moss B. 2012. Pervasive initiation and 3'-end formation of poxvirus postreplicative RNAs. *J Biol Chem* 287:31050–31060. <https://doi.org/10.1074/jbc.M112.390054>
17. Condit RC, Moussatche N, Traktman P. 2006. In a nutshell: structure and assembly of the vaccinia virion. *Adv Virus Res* 66:31–124. [https://doi.org/10.1016/S0065-3527\(06\)66002-8](https://doi.org/10.1016/S0065-3527(06)66002-8)
18. Tombác D, Prazsák I, Szucs A, Dénes B, Snyder M, Boldogkoi Z. 2018. Dynamic transcriptome profiling dataset of vaccinia virus obtained from long-read sequencing techniques. *Gigascience* 7:giy139. <https://doi.org/10.1093/gigascience/giy139>
19. McFadden G. 2005. Poxvirus tropism. *Nat Rev Microbiol* 3:201–213. <https://doi.org/10.1038/nrmicro1099>
20. Moss B. 2007. *Field's virology*. 5th ed. Lippincott Williams & Wilkins, Wolters Kluwer, Philadelphia.
21. Davison AJ, Moss B. 1989. Structure of vaccinia virus late promoters. *J Mol Biol* 210:771–784. [https://doi.org/10.1016/0022-2836\(89\)90108-3](https://doi.org/10.1016/0022-2836(89)90108-3)
22. Dhungel P, Cao S, Yang Z. 2017. The 5'-poly(A) leader of poxvirus mRNA confers a translational advantage that can be achieved in cells with impaired cap-dependent translation. *PLoS Pathog* 13:e1006602. <https://doi.org/10.1371/journal.ppat.1006602>
23. Alkhalil A, Hammamieh R, Hardick J, Ichou MA, Jett M, Ibrahim S. 2010. Gene expression profiling of monkeypox virus-infected cells reveals novel interfaces for host-virus interactions. *Virology* 403:173–183. <https://doi.org/10.1016/j.virol.2010.07.013>
24. Rubins KH, Hensley LE, Bell GW, Wang C, Lefkowitz EJ, Brown PO, Relman DA. 2008. Comparative analysis of viral gene expression programs during poxvirus infection: a transcriptional map of the



- vaccinia and monkeypox genomes. *PLoS One* 3:e2628. <https://doi.org/10.1371/journal.pone.0002628>
25. Rubins KH, Hensley LE, Relman DA, Brown PO. 2011. Stunned silence: gene expression programs in human cells infected with monkeypox or vaccinia virus. *PLoS ONE* 6:e15615. <https://doi.org/10.1371/journal.pone.0015615>
  26. Bourquain D, Dabrowski PW, Nitsche A. 2013. Comparison of host cell gene expression in cowpox, monkeypox or vaccinia virus-infected cells reveals virus-specific regulation of immune response genes. *Virology* 453:61–74. <https://doi.org/10.1016/j.virol.2013.08.011>
  27. Tombácz D, Prazsák I, Torma G, Csabai Z, Balázs Z, Moldován N, Dénes B, Snyder M, Boldogkői Z. 2021. Time-course transcriptome profiling of a poxvirus using long-read full-length assay. *Pathogens* 10:919. <https://doi.org/10.3390/pathogens10080919>
  28. Berk AJ, Sharp PA. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 12:721–732. [https://doi.org/10.1016/0092-8674\(77\)90272-0](https://doi.org/10.1016/0092-8674(77)90272-0)
  29. Weaver RF, Weissmann C. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S  $\beta$ -globin mRNA precursor and mature 10 S  $\beta$ -globin mRNA have identical map coordinates. *Nucleic Acids Res* 7:1175–1193. <https://doi.org/10.1093/nar/7.5.1175>
  30. Lee-Chen G-J, Niles EG. 1988. Map positions of the 5' ends of eight mRNAs synthesized from the late genes in the vaccinia virus HindIII D fragment. *Virology (Auckl)* 163:80–92. [https://doi.org/10.1016/0042-6822\(88\)90235-8](https://doi.org/10.1016/0042-6822(88)90235-8)
  31. Traktman P, Sridhar P, Condit RC, Roberts BE. 1984. Transcriptional mapping of the DNA polymerase gene of vaccinia virus. *J Virol* 49:125–131. <https://doi.org/10.1128/JVI.49.1.125-131.1984>
  32. Hruby DE, Maki RA, Miller DB, Ball LA. 1983. Fine structure analysis and nucleotide sequence of the vaccinia virus thymidine kinase gene. *Proc Natl Acad Sci USA* 80:3411–3415. <https://doi.org/10.1073/pnas.80.11.3411>
  33. Jones EV, Moss B. 1985. Transcriptional mapping of the vaccinia virus DNA polymerase gene. *J Virol* 53:312–315. <https://doi.org/10.1128/JVI.53.1.312-315.1985>
  34. Lutttge BG, Moyer RW. 2005. Suppressors of a host range mutation in the rabbitpox virus serpin SPI-1 map to proteins essential for viral DNA replication. *J Virol* 79:9168–9179. <https://doi.org/10.1128/JVI.79.14.9168-9179.2005>
  35. Lee-Chen G-J, Bourgeois N, Davidson K, Condit RC, Niles EG. 1988. Structure of the transcription initiation and termination sequences of seven early genes in the vaccinia virus HindIII D fragment. *Virology (Auckl)* 163:64–79. [https://doi.org/10.1016/0042-6822\(88\)90234-6](https://doi.org/10.1016/0042-6822(88)90234-6)
  36. Antczak JB, Patel DD, Ray CA, Ink BS, Pickup DJ. 1992. Site-specific RNA cleavage generates the 3' end of a poxvirus late mRNA. *Proc Natl Acad Sci USA* 89:12033–12037. <https://doi.org/10.1073/pnas.89.24.12033>
  37. D'Costa SM, Antczak JB, Pickup DJ, Condit RC. 2004. Post-transcription cleavage generates the 3' end of F17R transcripts in vaccinia virus. *Virology (Auckl)* 319:1–11. <https://doi.org/10.1016/j.virol.2003.09.041>
  38. Frohman MA, Dush MK, Martin GR. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85:8998–9002. <https://doi.org/10.1073/pnas.85.23.8998>
  39. Wennier ST, Brinkmann K, Steinhäuser C, Mayländer N, Mnich C, Wielert U, Dirmeier U, Hausmann J, Chaplin P, Steigerwald R. 2013. A novel naturally occurring tandem promoter in modified vaccinia virus ankara drives very early gene expression and potent immune responses. *PLoS One* 8:e73511. <https://doi.org/10.1371/journal.pone.0073511>
  40. Ensinger MJ, Martin SA, Paoletti E, Moss B. 1975. Modification of the 5'-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. *Proc Natl Acad Sci USA* 72:2525–2529. <https://doi.org/10.1073/pnas.72.7.2525>
  41. Parrish S, Moss B. 2007. Characterization of a second vaccinia virus mRNA-decapping enzyme conserved in poxviruses. *J Virol* 81:12973–12978. <https://doi.org/10.1128/JVI.01668-07>
  42. Parrish S, Resch W, Moss B. 2007. Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. *Proc Natl Acad Sci U S A* 104:2139–2144. <https://doi.org/10.1073/pnas.0611685104>
  43. Furuichi Y. 2015. Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc Jpn Acad Ser B Phys Biol Sci* 91:394–409. <https://doi.org/10.2183/pjab.91.394>
  44. Wei CM, Moss B. 1975. Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci USA* 72:318–322. <https://doi.org/10.1073/pnas.72.1.318>
  45. Kanagawa T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* 96:317–323. [https://doi.org/10.1016/S1389-1723\(03\)90130-7](https://doi.org/10.1016/S1389-1723(03)90130-7)
  46. Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics* 88:127–131. <https://doi.org/10.1016/j.ygeno.2005.12.013>
  47. Balázs Z, Tombácz D, Csabai Z, Moldován N, Snyder M, Boldogkői Z. 2019. Template-switching artifacts resemble alternative polyadenylation. *BMC Genomics* 20:824. <https://doi.org/10.1186/s12864-019-6199-7>
  48. Mourão K, Schurch NJ, Lucoszek R, Froussios K, MacKinnon K, Duc C, Simpson G, Barton GJ. 2019. Detection and mitigation of spurious antisense expression with RoSA. *F1000Res* 8:819. <https://doi.org/10.12688/f1000research.18952.1>
  49. McCarthy A. 2010. Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol* 17:675–676. <https://doi.org/10.1016/j.chembiol.2010.07.004>
  50. Thompson JF, Milos PM. 2011. The properties and applications of single-molecule DNA sequencing. *Genome Biol* 12:217. <https://doi.org/10.1186/gb-2011-12-2-217>
  51. Bayega A, Wang YC, Oikonomopoulos S, Djambazian H, Fahiminiya S, Ragoussis J. 2018. Transcript profiling using long-read sequencing technologies. *Methods Mol Biol* 1783:121–147. [https://doi.org/10.1007/978-1-4939-7834-2\\_6](https://doi.org/10.1007/978-1-4939-7834-2_6)
  52. Boldogkői Z, Moldován N, Balázs Z, Snyder M, Tombácz D. 2019. Long-read sequencing – a powerful tool in viral transcriptome research. *Trends Microbiol* 27:578–592. <https://doi.org/10.1016/j.tim.2019.01.010>
  53. Soneson C, Yao Y, Bratus-Neuenschwander A, Patrignani A, Robinson MD, Hussain S. 2019. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun* 10:3359. <https://doi.org/10.1038/s41467-019-11272-z>
  54. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA* 100:15776–15781. <https://doi.org/10.1073/pnas.2136655100>
  55. Carninci P, Kamada T, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C. 1996. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* 37:327–336. <https://doi.org/10.1006/geno.1996.0567>
  56. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. 2011. Genome-wide analysis of the 5' and 3' ends of vaccinia virus early mRNAs delineates regulatory sequences of annotated and anomalous transcripts. *J Virol* 85:5897–5909. <https://doi.org/10.1128/JVI.00428-11>
  57. Tombácz D, Prazsák I, Csabai Z, Moldován N, Dénes B, Snyder M, Boldogkői Z. 2020. Long-read assays shed new light on the transcriptome complexity of a viral pathogen. *Sci Rep* 10:13822. <https://doi.org/10.1038/s41598-020-70794-5>
  58. Maróti Z, Tombácz D, Prazsák I, Moldován N, Csabai Z, Torma G, Balázs Z, Kalmár T, Dénes B, Snyder M, Boldogkői Z. 2021. Time-course transcriptome analysis of host cell response to poxvirus infection using a dual long-read sequencing approach. *BMC Res Notes* 14:239. <https://doi.org/10.1186/s13104-021-05657-x>
  59. Wang Y, Zhang J, Li M, Jia M, Yang L, Wang T, Wang Y, Kang L, Li M, Kong L. 2024. Transcriptome and proteomic analysis of mpox virus F3L-expressing cells. *Front Cell Infect Microbiol* 14:1354410. <https://doi.org/10.3389/fcimb.2024.1354410>
  60. Desingu PA, Rubeni TP, Sundaresan NR. 2022. Evolution of monkeypox virus from 2017 to 2022: In the light of point mutations. *Front Microbiol* 13:1037598. <https://doi.org/10.3389/fmicb.2022.1037598>
  61. Kibungu EM, Vakaniaki EH, Kinganda-Lusamaki E, Kalonji-Mukendi T, Pukuta E, Hoff NA, Bogoch II, Cevik M, Gonsalves GS, Hensley LE, et al. 2024. Clade I-associated mpox cases associated with sexual contact, the



- Democratic Republic of the Congo. *Emerg Infect Dis* 30:172–176. <https://doi.org/10.3201/eid3001.231164>
62. Patel DD, Pickup DJ. 1987. Messenger RNAs of a strongly-expressed late gene of cowpox virus contain 5'-terminal poly(A) sequences. *EMBO J* 6:3787–3794. <https://doi.org/10.1002/j.1460-2075.1987.tb02714.x>
  63. Ahn BY, Moss B. 1989. Capped poly(A) leaders of variable lengths at the 5' ends of vaccinia virus late mRNAs. *J Virol* 63:226–232. <https://doi.org/10.1128/JVI.63.1.226-232.1989>
  64. Miner JN, Weinrich SL, Hruby DE. 1988. Molecular dissection of cis-acting regulatory elements from 5'-proximal regions of a vaccinia virus late gene cluster. *J Virol* 62:297–304. <https://doi.org/10.1128/JVI.62.1.297-304.1988>
  65. Cowley R, Greenaway PJ. 1990. Nucleotide sequence comparison of homologous genomic regions from variola, monkeypox, and vaccinia viruses. *J Med Virol* 31:267–271. <https://doi.org/10.1002/jmv.1890310405>
  66. Ahn BY, Gershon PD, Jones EV, Moss B. 1990. Identification of *rpo30*, a vaccinia virus RNA polymerase gene with structural similarity to a eucaryotic transcription elongation factor. *Mol Cell Biol* 10:5433–5441. <https://doi.org/10.1128/mcb.10.10.5433-5441.1990>
  67. Ferrareze PAG, Pereira e Costa RA, Thompson CE. 2023. Genomic characterization and molecular evolution of human monkeypox viruses. *Arch Virol* 168:278. <https://doi.org/10.1007/s00705-023-05904-5>
  68. Molteni C, Forni D, Cagliani R, Mozzi A, Clerici M, Sironi M. 2023. Evolution of the orthopoxvirus core genome. *Virus Res* 323:198975. <https://doi.org/10.1016/j.virusres.2022.198975>
  69. Rosel JL, Earl PL, Weir JP, Moss B. 1986. Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the HindIII H genome fragment. *J Virol* 60:436–449. <https://doi.org/10.1128/JVI.60.2.436-449.1986>
  70. Yang Z, Reynolds SE, Martens CA, Bruno DP, Porcella SF, Moss B. 2011. Expression profiling of the intermediate and late stages of poxvirus replication. *J Virol* 85:9899–9908. <https://doi.org/10.1128/JVI.05446-11>
  71. Balázs Z, Tombácz D, Szűcs A, Csabai Z, Megyeri K, Petrov AN, Snyder M, Boldogkői Z. 2017. Long-read sequencing of human cytomegalovirus transcriptome reveals RNA isoforms carrying distinct coding potentials. *Sci Rep* 7:15989. <https://doi.org/10.1038/s41598-017-16262-z>
  72. Mohamed MR, Niles EG. 2003. UUUUUNU stimulation of vaccinia virus early gene transcription termination. *J Biol Chem* 278:39534–39541. <https://doi.org/10.1074/jbc.M306048200>
  73. Ma Y, Chen M, Bao Y, Song S, MPoxVR Team. 2022. MPoxVR: a comprehensive genomic resource for monkeypox virus variant surveillance. *Innovation (Camb)* 3:100296. <https://doi.org/10.1016/j.xinn.2022.100296>
  74. Calvo-Roitberg E, Daniels RF, Pai AA. 2023. Challenges in identifying mRNA transcript starts and ends from long-read sequencing data. *bioRxiv*:2023.07.26.550536. <https://doi.org/10.1101/2023.07.26.550536>
  75. Moldován N, Torma G, Gulyás G, Hornyák Á, Zádori Z, Jefferson VA, Csabai Z, Boldogkői M, Tombácz D, Meyer F, Boldogkői Z. 2020. Time-course profiling of bovine alphaherpesvirus 1.1 transcriptome using multiplatform sequencing. *Sci Rep* 10:20496. <https://doi.org/10.1038/s41598-020-77520-1>
  76. Torma G, Tombácz D, Csabai Z, Moldován N, Mészáros I, Zádori Z, Boldogkői Z. 2021. Combined short and long-read sequencing reveals a complex transcriptomic architecture of African swine fever virus. *Viruses* 13:579. <https://doi.org/10.3390/v13040579>
  77. Trotman JB, Schoenberg DR. 2019. A recap of RNA recapping. *Wiley Interdiscip Rev RNA* 10:e1504. <https://doi.org/10.1002/wrna.1504>
  78. O'Grady T, Wang X, Höner Zu Bentrup K, Baddoo M, Concha M, Flemington EK. 2016. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res* 44:e145. <https://doi.org/10.1093/nar/gkw629>
  79. Braspenning SE, Sadaoka T, Breuer J, Verjans G, Ouwendijk WJD, Depledge DP. 2020. Decoding the architecture of the Varicella-Zoster virus transcriptome. *MBio* 11:e01568. <https://doi.org/10.1128/mBio.01568-20>
  80. Ibrahim F, Oppelt J, Maragkakis M, Mourelatos Z. 2021. TERA-Seq: true end-to-end sequencing of native RNA molecules for transcriptome characterization. *Nucleic Acids Res* 49:e115–e115. <https://doi.org/10.1093/nar/gkab713>
  81. Gubser C, Hué S, Kellam P, Smith GL. 2004. Poxvirus genomes: a phylogenetic analysis. *J Gen Virol* 85:105–117. <https://doi.org/10.1099/vir.0.19565-0>
  82. Kluczyk A, Siemion IZ, Szewczuk Z, Wieczorek Z. 2002. The immunosuppressive activity of peptide fragments of vaccinia virus C10L protein and a hypothesis on the role of this protein in the viral invasion. *Peptides* 23:823–834. [https://doi.org/10.1016/S0196-9781\(02\)00006-2](https://doi.org/10.1016/S0196-9781(02)00006-2)
  83. Herbert MH, Squire CJ, Mercer AA. 2015. Poxviral ankyrin proteins. *Viruses* 7:709–738. <https://doi.org/10.3390/v7020709>
  84. Yang Z, Cao S, Martens CA, Porcella SF, Xie Z, Ma M, Shen B, Moss B. 2015. Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling. *J Virol* 89:6874–6886. <https://doi.org/10.1128/JVI.00528-15>
  85. Boyle KA, Greseth MD, Traktman P. 2015. Genetic confirmation that the H5 protein is required for vaccinia virus DNA replication. *J Virol* 89:6312–6327. <https://doi.org/10.1128/JVI.00445-15>
  86. Rabaan AA, Abas AH, Tallei TE, Al - Zaher MA, Al - Sheef NM, Fatimawali, Al - Nass EZ, Al - Ebrahim EA, Effendi Y, Idroes R, Alhabib MF, Al - Fheid HA, Adam AA, Bin Emran T. 2023. Monkeypox outbreak 2022: what we know so far and its potential drug targets and management strategies. *J Med Virol* 95:e28306. <https://doi.org/10.1002/jmv.28306>
  87. Rach EA, Yuan H-Y, Majoros WH, Tomancak P, Ohler U. 2009. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the *Drosophila* genome. *Genome Biol* 10:R73. <https://doi.org/10.1186/gb-2009-10-7-r73>
  88. Haberle V, Forrest ARR, Hayashizaki Y, Carninci P, Lenhard B. 2015. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res* 43:e51. <https://doi.org/10.1093/nar/gkv054>
  89. Christen LA, Piacente S, Mohamed MR, Niles EG. 2008. Vaccinia virus early gene transcription termination factors VTF and Rap94 interact with the U9 termination motif in the nascent RNA in a transcription ternary complex. *Virology (Auckl)* 376:225–235. <https://doi.org/10.1016/j.virol.2008.03.031>
  90. Yang Z, Moss B. 2009. Interaction of the vaccinia virus RNA polymerase-associated 94-kilodalton protein with the early transcription factor. *J Virol* 83:12018–12026. <https://doi.org/10.1128/JVI.01653-09>
  91. Vopálenský V, Sýkora M, Mělková Z, Mašek T, Pospíšek M. 2020. Transcripts of vaccinia virus postreplicative genes do not contain a 5' methylguanosine cap. *bioRxiv*. <https://doi.org/10.1101/2020.07.15.204867>
  92. Farlow J, Ichou MA, Huggins J, Ibrahim S. 2010. Comparative whole genome sequence analysis of wild-type and cidofovir-resistant monkeypoxvirus. *Virol J* 7:110. <https://doi.org/10.1186/1743-422X-7-110>
  93. Scarpa F, Sanna D, Azzena I, Cossu P, Locci C, Angeletti S, Maruotti A, Ceccarelli G, Casu M, Fiori PL, Petrosillo N, Ciccozzi M. 2022. Genetic variability of the monkeypox virus clade IIb B.1. *J Clin Med* 11:6388. <https://doi.org/10.3390/jcm11216388>
  94. Ndodo N, Ashcroft J, Lewandowski K, Yinka-Ogunleye A, Chukwu C, Ahmad A, King D, Akinpelu A, Maluquer de Motes C, Ribeca P, Sumner RP, Rambaut A, Chester M, Maishman T, Bamidele O, Mba N, Babatunde O, Aruna O, Pullan ST, Gannon B, Brown CS, Ihekweazu C, Adetifa I, Ulaeto DO. 2023. Distinct monkeypox virus lineages co-circulating in humans before 2022. *Nat Med* 29:2317–2324. <https://doi.org/10.1038/s41591-023-02456-8>
  95. Suspène R, Raymond KA, Boutin L, Guillier S, Lemoine F, Ferraris O, Tournier J-N, Iseni F, Simon-Lorière E, Vartanian J-P. 2023. APOBEC3F is a mutational driver of the human monkeypox virus identified in the 2022 outbreak. *J Infect Dis* 228:1421–1429. <https://doi.org/10.1093/infdis/jiad165>
  96. Delamonica B, Davalos L, Larijani M, Anthony SJ, Liu J, MacCarthy T. 2023. Evolutionary potential of the monkeypox genome arising from interactions with human APOBEC3 enzymes. *Virus Evol* 9:vead047. <https://doi.org/10.1093/ve/vead047>
  97. Jahankir MJB, Sounderrajan V, Rao SS, Thangam T, Kamariah N, Kurumbati A, Harshavardhan S, Ashwin A, Jeyaraj S, Parthasarathy K. 2024. Accelerated mutation by host protein APOBEC in monkeypox

- virus. *Gene Rep* 34:101878. <https://doi.org/10.1016/j.genrep.2024.101878>
98. Pfaff F, Hoffmann D, Beer M. 2022. Monkeypox genomic surveillance will challenge lessons learned from SARS-CoV-2. *The Lancet* 400:22–23. [https://doi.org/10.1016/S0140-6736\(22\)01106-0](https://doi.org/10.1016/S0140-6736(22)01106-0)
  99. Gigante CM, Korber B, Seabolt MH, Wilkins K, Davidson W, Rao AK, Zhao H, Smith TG, Hughes CM, Minhaj F, et al. 2022. Multiple lineages of monkeypox virus detected in the United States, 2021–2022. *Science* 378:560–565. <https://doi.org/10.1126/science.add4153>
  100. Lopera JG, Falendysz EA, Rocke TE, Osorio JE. 2015. Attenuation of monkeypox virus by deletion of genomic regions. *Virology (Auckl)* 475:129–138. <https://doi.org/10.1016/j.virol.2014.11.009>
  101. Li H, Zhang H, Ding K, Wang X-H, Sun G-Y, Liu Z-X, Luo Y. 2022. The evolving epidemiology of monkeypox virus. *Cytokine Growth Factor Rev* 68:1–12. <https://doi.org/10.1016/j.cytogfr.2022.10.002>
  102. Baroudy BM, Moss B. 1982. Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. *Nucleic Acids Res* 10:5673–5679. <https://doi.org/10.1093/nar/10.18.5673>
  103. Qin L, Upton C, Hazes B, Evans DH. 2011. Genomic analysis of the vaccinia virus strain variants found in Dryvax vaccine. *J Virol* 85:13049–13060. <https://doi.org/10.1128/JVI.05779-11>
  104. Tulman ER, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. 2004. The genome of canarypox virus. *J Virol* 78:353–366. <https://doi.org/10.1128/jvi.78.1.353-366.2004>
  105. Desingu PA, Nagarajan K, Sundaresan NR. 2023. Unique tandem repeats in the inverted terminal repeat regions of monkeypox viruses. *Microbiol Spectr* 11:e0319922. <https://doi.org/10.1128/spectrum.03199-22>
  106. Desingu PA, Nagarajan K. 2022. Genomic regions insertion and deletion in monkeypox virus causing multi-country outbreak-2022. *bioRxiv*. <https://doi.org/10.1101/2022.06.28.497936>
  107. Lu Z, Berry K, Hu Z, Zhan Y, Ahn T-H, Lin Z. 2021. TSSr: an R package for comprehensive analyses of TSS sequencing data. *NAR Genom Bioinform* 3:lqab108. <https://doi.org/10.1093/nargab/lqab108>
  108. Wyman D, Balderrama-Gutierrez G, Reese F, Jiang S, Rahmanian S, Forner S, Matheos D, Zeng W, Williams B, Trout D, England W, Chu S-H, Spitale RC, Tenner AJ, Wold BJ, Mortazavi A. 2020. A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. *bioRxiv*. <https://doi.org/10.1101/672931>
  109. Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27:1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>
  110. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, Zuzarte PC, Gilpatrick T, Payne A, Quick J, Sadowski N, Holmes N, de Jesus JG, Jones KL, Soulette CM, Snutch TP, Loman N, Paten B, Loose M, Simpson JT, Olsen HE, Brooks AN, Akeson M, Timp W. 2019. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* 16:1297–1305. <https://doi.org/10.1038/s41592-019-0617-2>
  111. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
  112. Desingu PA, Rubeni TP, Nagarajan K, Sundaresan NR. 2024. Molecular evolution of 2022 multi-country outbreak-causing monkeypox virus Clade IIb. *i Sci* 27:108601. <https://doi.org/10.1016/j.isci.2023.108601>