Analysis of MPXV RNA-seq Data Reveals Lack of Evidence of APOBEC3-mediated RNA Editing

Alisa O. Lyskova [1,*], Ruslan Kh. Abasov [1,2,*], Anna Pavlova [1], Evgenii V. Matveev [1,2], Alexandra Madorskaya [3], Fedor M. Kazanov [3], Gennady V. Ponomarev [2], Dmitry N. Ivankov [4], Ogun Adebali [5], Mikhail S. Gelfand [4], Marat D. Kazanov [1,2,4,5]

[1] Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia.
[2] Vavilov Institute of General Genetics, Moscow, Russia.
[3] National Research University Higher School of Economics, Moscow, Russia.
[4] Center of Molecular and Cellular Biology, Moscow, Russia.
[5] Sabanci University, Istanbul, Türkiye.
[*] equal contribution
Correspondence: mkazanov@gmail.com

Abstract: The 2022 outbreak of monkeypox virus (MPXV), a double-stranded DNA virus, is remarkable for an unusually high number of single-nucleotide substitutions compared to earlier strains, with a strong bias toward C→T and G→A transitions consistent with the APOBEC3 cytidine deaminase activity. While APOBEC3-induced mutagenesis is well documented at the DNA level, its potential impact on MPXV RNA transcripts remains unclear. To assess whether APOBEC3 enzymes act on MPXV RNA, we analyzed RNA-seq data from infected samples. The enrichment of APOBEC-signature substitutions among high-frequency mismatched positions led us to consider two possibilities: RNA editing at hotspots or fixed DNA mutations. Multiple lines of evidence support the conclusion that these substitutions arise from DNA-level mutagenesis rather than RNA editing. These include a substantial number of G→A substitutions remaining after normalization by gene strand direction, a largely neutral impact of substitutions on protein-coding sequences, the lack of positional correlation with transcriptional features or RNA secondary structure typically associated with APOBEC action hotspots, and an overlap with known genomic mutations in MPXV strains. Analysis of the nucleotide context of observed substitutions indicated that APOBEC3A was a likely driver of DNA-level mutagenesis.

Introduction

Monkeypox disease, known for a multi-country outbreak in 2022, is caused by the monkeypox virus (MPXV), a double-stranded DNA virus belonging to the Orthopoxvirus genus in the Poxviridae family[1]. MPXV was first isolated in 1958 in Copenhagen from a *Macaca fascicularis* imported from Singapore [2]. The first human case of MPXV was documented in 1970 in the Congo Basin [3], demonstrating the virus's capability for a zoonotic transmission from animals to humans. Since then, numerous outbreaks were reported across Eastern, Central, and Western Africa [4]. The symptoms of monkeypox are similar to those of smallpox, which also belongs to the Orthopoxvirus genus along with cowpox, camelpox, mousepox, and vaccinia viruses [5]. However, monkeypox typically has a lower mortality rate [6]. The natural reservoir of MPXV is thought to be rodents and other small animals [7]. From 2002 to 2021, several cases of MPXV reported outside of Africa were associated with imports from endemic countries in West and Central Africa [8,9]. In 2022, a rapidly expanding multi-country outbreak affected individuals with no travel history to Africa, suggesting that MPXV may have evolved to enable human-to-human transmission [10]. Since the start of the outbreak, more than 90,000 cases were reported across over 110 countries, leading the World Health Organization (WHO) to declare it a global health emergency.

MPXV is phylogenetically divided into three distinct clades: Clade I, the most virulent one, leads to up to 10% mortality in humans and is predominantly transmitted in the Congo Basin; Clade IIa, initially a zoonotic strain with low mortality in West Africa, later adapted to the human-to-human transmission and caused an outbreak in Nigeria during 2017–2018; and Clade IIb, associated with the 2022 global outbreak, characterized by its ability to spread through human-to-human transmission [11,12]. The MPXV genome

is approximately 197,000 base pairs long and contains about 200 genes [13]. A relatively large number of genes are involved in to the encoded replication machinery, as the entire viral lifecycle takes place in the cytoplasm of infected cells [14]. The MPXV genome includes two telomeres with identical sequences that are oriented in opposite directions, known as inverted terminal repeats (ITRs) [15]. Genes responsible for replication and morphogenesis, conserved among poxviruses, are clustered in the central region of the genome, while the genes specific to pathogenesis are located in the terminal regions [16]. Studies on the transcriptional program of MPXV categorized its genes into three temporal expression phases: early, intermediate, and late [17]. The estimated mutation rate for orthopoxviruses is 1–2 substitutions per genome per year [18], however, the extent of divergence in individual genomes from the 2022 outbreak, compared to the related 2018–2019 viruses, has reached approximately 50 single-nucleotide substitutions [12]. Most substitutions were C→T transitions or their complementary G→A counterparts, typically occurring in the TC (complementary GA) contexts, respectively [12]. This mutational signature, characteristic of APOBEC3 cytidine deaminases, suggests that the accumulation of these mutations in the MPXV genome could indicate the virus's evolved ability for human-to-human transmission [12,19,20].

Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) is a family of enzymes in mammals that play a key role in the immune system, restricting viruses and retrotransposons [21]. Recently, this family of enzymes has received significant attention for its role in the cancer mutagenesis and its potential involvement in the mutagenesis of the SARS-CoV-2 virus [22–24]. APOBEC enzymes bind to RNA or single-stranded DNA, catalyzing deamination of cytosine to uracil. In an RNA molecule, the altered nucleotide will remain unchanged. In a DNA molecule, uracil can be fixed as thymine in subsequent rounds of replication or, if converted by DNA glycosylases into an abasic site, it may be further repaired by translesion polymerases into either thymine or guanine [25]. It has been observed that there exist mutagenesis hotspots of APOBEC enzymes occurring within particular secondary structure patterns of RNA or single-stranded DNA [26]. Specifically, a cytosine located at the 3' end of a hairpin loop would have a 10-fold higher mutation rate by APOBEC enzymes compared to cytosines in TC motifs outside secondary structure [27]. A similar effect was observed for APOBEC signature mutations in the MPXV genome, providing additional support for the involvement of APOBEC enzymes [28]. While mutations in MPXV genomes associated with the 2022 outbreak have been extensively studied, it remains unclear whether viral messenger RNAs are also affected by APOBEC enzymes.

In this study, we analyzed publicly available MPXV RNA-seq data for evidence of APOBEC mutagenesis in viral transcripts. Analysis of MPXV transcriptomes generated by RNA-sequencing can provide valuable insights into the extent and characteristics of APOBEC mutagenesis. Moreover, RNA-seq studies can generate significantly more statistical data than genome sequencing, as messenger RNAs are produced in multiple copies during gene expression, and each of these molecules can theoretically be targeted by APOBEC enzymes. Our results reveal a moderate number of APOBEC-signature substitutions in MPXV transcripts which, upon thorough analysis, appear to stem from DNA-level mutagenesis, hence limiting support for APOBEC-mediated RNA editing.

Methods

RNA-seq datasets from the following projects were obtained from the SRA archive: PRJEB60728 (host: *Homo sapiens*), PRJNA906618 (host: *Homo sapiens*), PRJEB56841 (host: *Chlorocebus sabaeus*), PRJNA1183318 (host: *Macaca fascicularis*) along with DNA sequencing data from projects PRJNA845087 and PRJNA981509 (both host: *Homo sapiens*). Both RNA-seq and DNA-seq datasets were downloaded using the fastq-dump utility from the SRA Toolkit (version 3.1.1) [29]. The reference genome of the monkeypox virus (MPXV) (GCF_014621545.1 / ASM1462154v1), as well as host genomes (*Homo sapiens:* GRCh38.d1.v1, *Chlorocebus sabaeus:* GCF_015252025.1, *Macaca fascicularis:* GCA_037993035.1), were obtained from the NCBI RefSeq database [30].

RNA-seq data generated using long-read sequencing technologies were aligned to the reference genome using minimap2 (version 2.28-r1209) [31]. Short-read RNA-seq data were aligned using STAR (version 2.7.11b) [32], and genomic DNA short reads were mapped using BWA (version 2.2.1) [33]. Alignment quality metrics were computed using mosdepth (version 0.3.9) [34] and the samtools stats command from samtools (version 1.20) [35]. Detection of positions with mismatches, representing either genomic variants or potential RNA editing sites, was performed using BCFtools (version 1.21, the exact command-line parameters are provided in Table S1) [35]. Mismatch positions were filtered using a substitution frequency

threshold of 1%, which is ten times higher than the reported sequencing error rate [36].

Secondary structure around the identified mismatch positions was predicted using RNAfold [37] and further analyzed with the RNAsselem package (version 1.3) [38]. Inverted repeats were identified using Palindrome Analyzer [39]. The genomic map was visualized using the circlize package in R. All other plots were generated using ggplot2 (version 3.5.1) and matplotlib (version 3.9.0).

To model the sequence specificity of APOBEC family enzymes (APOBEC3A, APOBEC3B, APOBEC3F, and APOBEC3G), we constructed position weight matrices (PWMs) based on published data reporting experimentally confirmed mutation sites and their sequence contexts [40–42]. For each enzyme, we compiled nucleotide sequences flanking known editing sites from the literature and counted the frequency of each nucleotide (A, C, G, T) at each position within a defined 5-nucleotide window (−3, −2, −1, 0, 1) surrounding the mutation site, where position 0 corresponds to the mutated nucleotide. These nucleotide counts were assembled into a matrix, with each row corresponding to a nucleotide and each column to a relative position. The counts in each column were normalized to sum to 1, and the resulting frequencies were log-transformed to generate the final PWM. To score a given sequence position, we summed the logarithmic weights corresponding to the observed nucleotide at each position in the matrix. Additionally, PWMs specifically designed to distinguish APOBEC3A from APOBEC3B were adopted from [43], based on the differential preference for YTCA versus RTCA sequence motifs (the mutated C is underlined).

Results

APOBEC-signature substitutions in MPXV transcripts are observed at a limited number of positions, where they occur at high frequency

To identify RNA editing events in MPXV transcripts, we comprehensively analyzed available RNA-seq data from infected hosts, including two human [44,45], and two macaque studies [46,47] (hereafter referred to as RNAseq:human1–2 and RNAseq:monkey1–2). To ensure accurate detection of potential RNA editing events, we mapped sequenced reads to the MPXV genome using three complementary approaches: (i) direct mapping to the MPXV genome, (ii) mapping to the combined host and MPXV genome, and (iii) initial mapping to the host genome, followed by mapping of the unmapped reads to the MPXV genome. Following read mapping, we identified positions with mismatches using BCFtools allowing for highly sensitive detection of substitutions. To achieve maximum specificity, we set the threshold for substitution frequency as low as possible, but substantially higher than the probability of sequencing errors (see Methods). We applied this method to three alignments generated for each RNA-seq study and derived the consensus positions (Fig. 1a).

Subsequent analysis of the detected transcript positions revealed similar results across the examined projects (Supplemental file 1), with Figure 1 illustrating a representative example using a single sample ERR11030177 from project [44]. In this sample, more than 10,000 positions with potential RNA editing events were identified using the BCFtools method. We analyzed the fractions of substitution types at the detected positions that could be attributed to APOBEC-signature substitutions, specifically C→T or complementary G→A changes in the TC/GA context. Alternatively, we considered the same substitutions occurring outside of this context as part of the extended APOBEC signature [48]. We observed a moderate enrichment of C→T and complementary G→A substitutions (Fig. 1b) and analyzed their two-nucleotide context, revealing that a significant fraction of these substitutions had occurred within the APOBEC TC/GA motif (Fig. 1b, 1c). However, the analysis of distribution of substitution frequencies (SF) at detected positions (Fig. 1d) revealed that most APOBEC-signature positions have SF greater than 50%, whereas positions with medium and low SF show no enrichment of APOBEC-signature substitutions (Fig. 1e). We assumed that the high frequency of APOBEC-signature substitutions at specific transcript positions could result from either pre-existing DNA-level substitutions or RNA editing hotspots within transcripts. To distinguish between these alternatives, we conducted an additional analysis related to the identified positions. For subsequent analysis, we grouped substitution positions detected in samples by the considered studies and filtered them by substitution frequency (SF > 0.5) to retain the range enriched for APOBEC-signature substitutions (Supplemental file 2).

Gene-focused analysis at positions of APOBEC-signature substitutions provides evidence that a significant fraction of substitutions occurred at the DNA level

When APOBEC-signature RNA editing occurs in a transcript, non-stranded RNA sequencing can gen-

Figure 1. (**a**) *Scheme of the pipeline for detection of variant-like positions in RNA-seq data, which may reflect either genomic variants or RNA editing events.* (**b**) *Distribution of observed substitutions in MPXV transcripts over substitution types.* (**c**) *DDistribution of observed substitutions in MPXV transcripts over three-nucleotide sequence context.* (**d**) *APOBEC-signature substitutions are enriched among positions with high mismatch frequency (>0.5).* (**e**) *Low-frequency (<0.5) mismatch sites are not enriched with APOBEC-like substitutions.*

erate both direct and complementary reads, yielding both C→T and G→A substitutions in the TC/GA context. However, considering gene direction, we can assess whether suspected RNA editing leads to C→T substitutions in genes on the plus strand and G→A substitutions in genes on the minus strand of the double-stranded MPXV DNA virus. We normalized all substitutions according to the gene strand direction, preserving those in plus-strand gene regions and converting those in minus-strand gene regions to their complementary counterparts. Following this normalization, the number of G→A substitutions decreased but remained substantial (Fig. 2a). We conclude that a likely explanation for this is that a significant fraction of substitutions occurred not in RNA transcripts, but at the DNA level before transcription.

Additional support for this hypothesis could come from estimating the effect of APOBEC-signature substitutions on the encoded amino acid sequence of transcripts. If these substitutions occur at the DNA level, deleterious substitutions may severely disturb the viral life cycle. Hence, most observed substitutions should be neutral or at worst mildly deleterious. We classified APOBEC-signature substitutions based on their impact on gene function and observed a depletion of non-synonymous substitutions, which are often deleterious, along with an enrichment of synonymous ones, which are generally considered neutral (Fig. 2b).

APOBEC-signature substitutions do not correlate with MPXV genomic features associated with transcription

To further distinguish between two hypotheses about the observed APOBEC signature substitutions – whether they result from RNA editing at transcript hotspots or from DNA-level substitutions – we analyzed the correlation between the identified genomic positions and known transcription-associated genomic features. If RNA editing predominates, we may expect the detected positions to correlate with transcriptional features, such as gene expression levels [49] or promoter regions. We analyzed the distribution of the identified positions along the MPXV genome (Fig.3) and found that APOBEC-signature substitutions, consistently detected by all three applied methods, were relatively uniformly distributed across the genome (Kolmogorov-Smirnov test: human, p=0.44; non-human primates, p=0.23). Then, we analyzed potential correlations between the positions and known gene categories classified according to gene expression timing. We used a well-established categorization of MPXV genes into early, intermediate and late transcription [50] but found no statistically significant difference in the density of positions with APOBEC-signature substitutions among these gene categories (Chi-squared test: human, $\chi 2$=3.3, p=0.5; non-human primates,

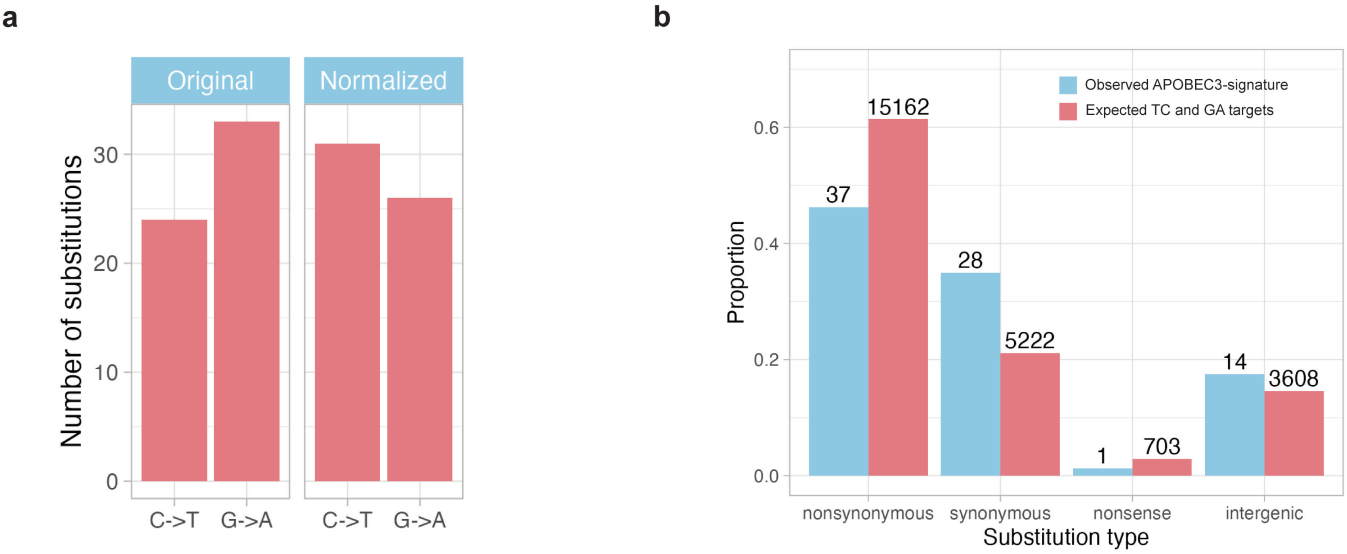

Figure 2. (**a**) *TThe number of G→A substitutions with SF > 0.5 identified in RNA-seq:human1-2 dataset remains elevated after normalization of C→T and G→A substitutions based on gene strand orientation.* (**b**) *Analysis of the effect of APOBEC-signature substitutions identified in RNA-seq:human1-2 dataset on the amino acid content encoded in MPXV transcripts reveals a prevalence of synonymous substitutions and a depletion of non-synonymous and nonsense substitutions.*

$\chi 2$=4.09, p=0.39). Among other genomic features associated with transcription in MPXV, we examined potential correlations with the positions of transcription start sites (TSS), transcription end sites (TES), and promoters, but found no statistically significant associations (human: TSS, p=0.37; TES, p=0.31; promoters, p=0.13; non-human primates: TSS, p=0.28; TES, p=0.37; promoters, p=0.11). Thus, our analysis of the genomic distribution of APOBEC-signature substitutions relative to transcription-linked genomic features found no significant associations.

Secondary structure at APOBEC-signature substitution sites lacks known APOBEC-associated hotspot patterns

Previous studies have reported that several members of the APOBEC family preferentially mutate TC motifs within specific secondary structure elements [26].

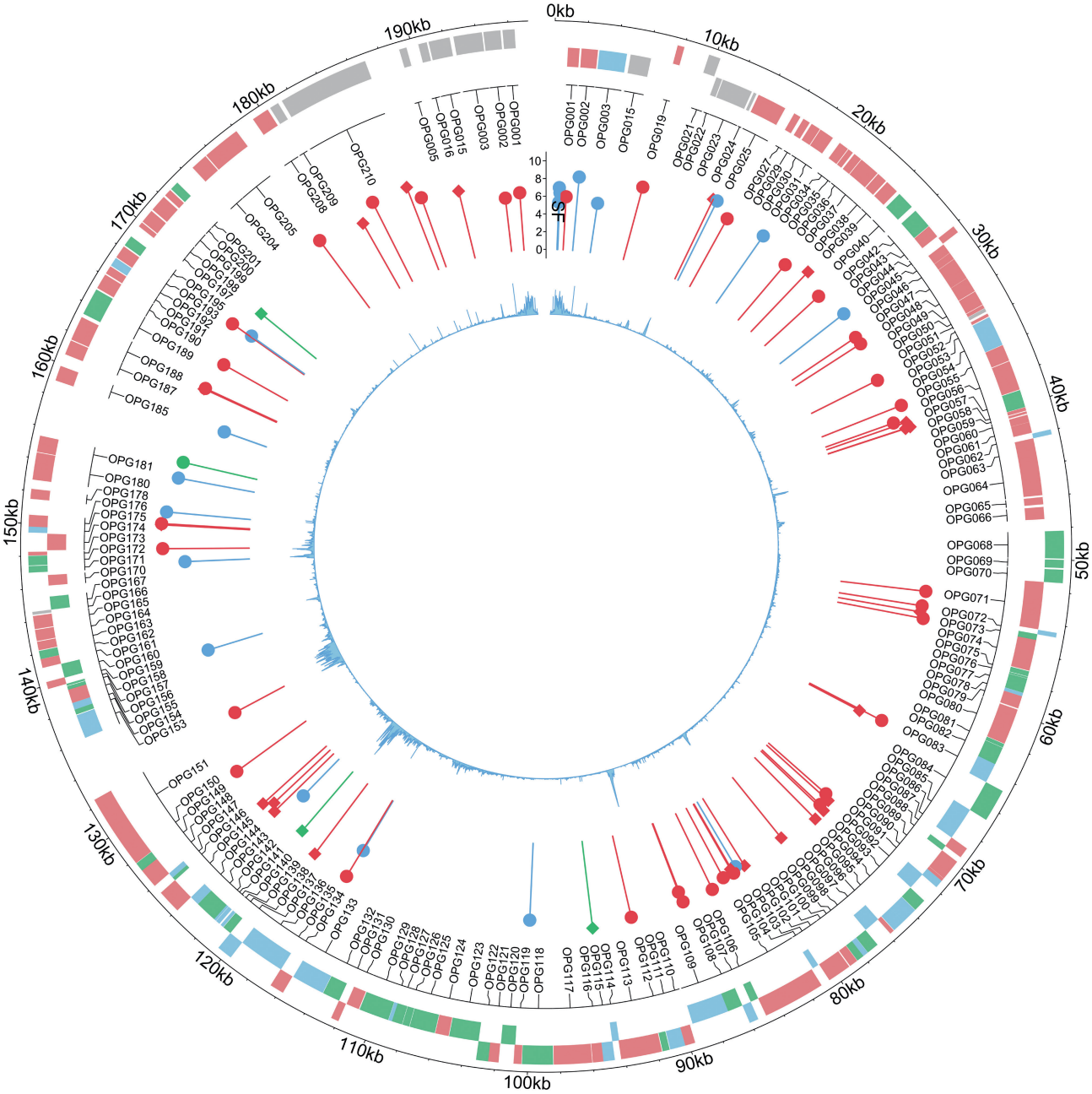

Figure 3. *Genomic map of identified substitution positions in the MPXV genome from two human-derived projects. The outermost track shows annotated viral genes, color-coded by expression phase: early (green), middle (red), and late (blue). The next layer inward displays the positions of identified substitutions. The height of each pin reflects the substitution frequency (SF). Pin color indicates substitution type: APOBEC-signature (red), extended APOBEC-signature (green), and other substitutions (blue). The pin head shape denotes strand orientation: a circle indicates a C→T substitution on the gene strand, while a square indicates the same substitution on the complementary strand. The innermost track displays the mean coverage across the genome.*

An elevated frequency of APOBEC-induced mutations was observed in cytosines located at the 3' end of hairpin loops, which can form in single-stranded DNA or RNA. To analyze whether the genomic positions with APOBEC-signature substitutions detected in our study are also associated with this secondary structure pattern, we applied the bioinformatics tool RNAsselem, which had been previously developed by our group [38]. We extended the functionality of RNAsselem, which was initially developed for analyzing experimentally known secondary structures of RNA viruses, to support predicted secondary structures. Using RNAsselem, we analyzed elements of the predicted secondary structure at positions with APOBEC-signature substitutions and compared them to predicted secondary structure patterns at all other positions in the viral genome. We found no association of APOBEC-signature substitution sites with hairpin loop structures, nor did we observe any statistically significant differences between secondary structures at APOBEC-signature positions and at other positions in the MPXV genome (Fig. 4a).

It was recently reported that mutations observed during the 2022 MPXV outbreak are enriched in inverted repeat (IR) regions [51]. However, our analysis of APOBEC-signature substitution sites within

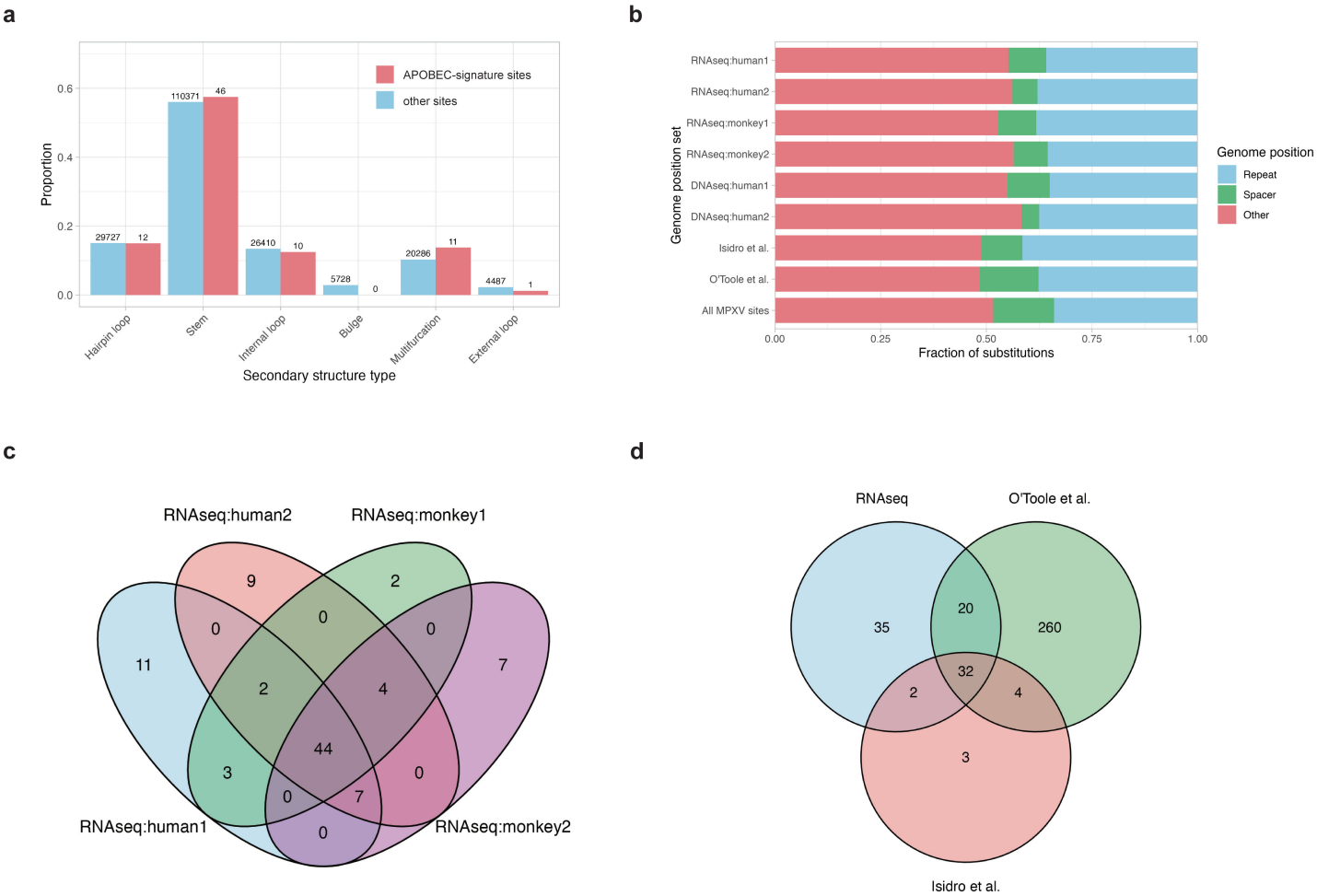

Figure 4. (**a**) *Distribution of predicted secondary structure elements at the positions of observed substitutions in RNA-seq:human1-2 dataset and at all other genomic positions. No statistically significant enrichment of substitutions is observed in hairpin loops – structures previously associated with APOBEC mutational hotspots.* (**b**) *Distribution of APOBEC-signature substitutions identified in the datasets analyzed in this study and in previous works, mapped onto inverted repeat elements in the MPXV genome that are potentially capable of forming hairpin structures. No significant difference is observed between substitution sites and the rest of the genome.* (**c**) *Venn diagram showing the overlap of APOBEC-signature substitutions identified in human and monkey RNA-seq datasets analyzed in this study.* (**d**) *Venn diagram showing the intersection of APOBEC-signature substitutions detected across all RNA-seq and DNA-seq datasets analyzed in this study and mutations reported in MPXV strains from O'Toole et al. and Isidro et al.*

hairpin structures – which inherently include IR regions – contradicts this finding. To clarify this discrepancy, we extended our RNA-seq dataset from human and non-human primate hosts by including two additional mutation sets obtained from DNA sequencing data presented in [52] (hereafter referred to as DNA-seq:human1–2), as well as two published MPXV mutation datasets from the 2022 outbreak [53,54]. We then applied the Palindrome Analyser tool used in [51]. Our results (Fig. 4b) showed no statistically significant enrichment of APOBEC-signature substitution sites in IR regions of the MPXV genome in either the RNA-seq or DNA-seq datasets (Chi-square test: RNA-seq:human1, p=0.45; RNA-seq:human2, p=0.16; RNA-seq:monkey1, p=0.5; RNA-seq:monkey2, p=0.36; DNA-seq:human1, p=0.62; DNA-seq:human2, p=0.05; Isidro et al. [53], p=0.5; O'Toole et al. [54], p=0.38). Moreover, we did not observe any trend suggesting such enrichment.

MPXV genome positions with APOBEC-signature substitutions significantly overlap across studies, reflecting pre-existing mutations rather than RNA editing hotspots

We also compared the lists of positions with APOBEC-signature substitutions identified in the analyzed RNA-seq datasets, DNA sequencing studies and known MPXV mutation datasets [53,54]. A substantial overlap of substitution sites was observed across different datasets (Fig. 4c). Specifically, 53 of 80 sites (66%) overlapped between two human RNA-seq studies, 48 of 69 sites (70%) between two non-human primate studies, and 44 out of 89 sites (49%) were common across all four RNA-seq datasets. Comparison with the comprehensive catalog of MPXV mutations

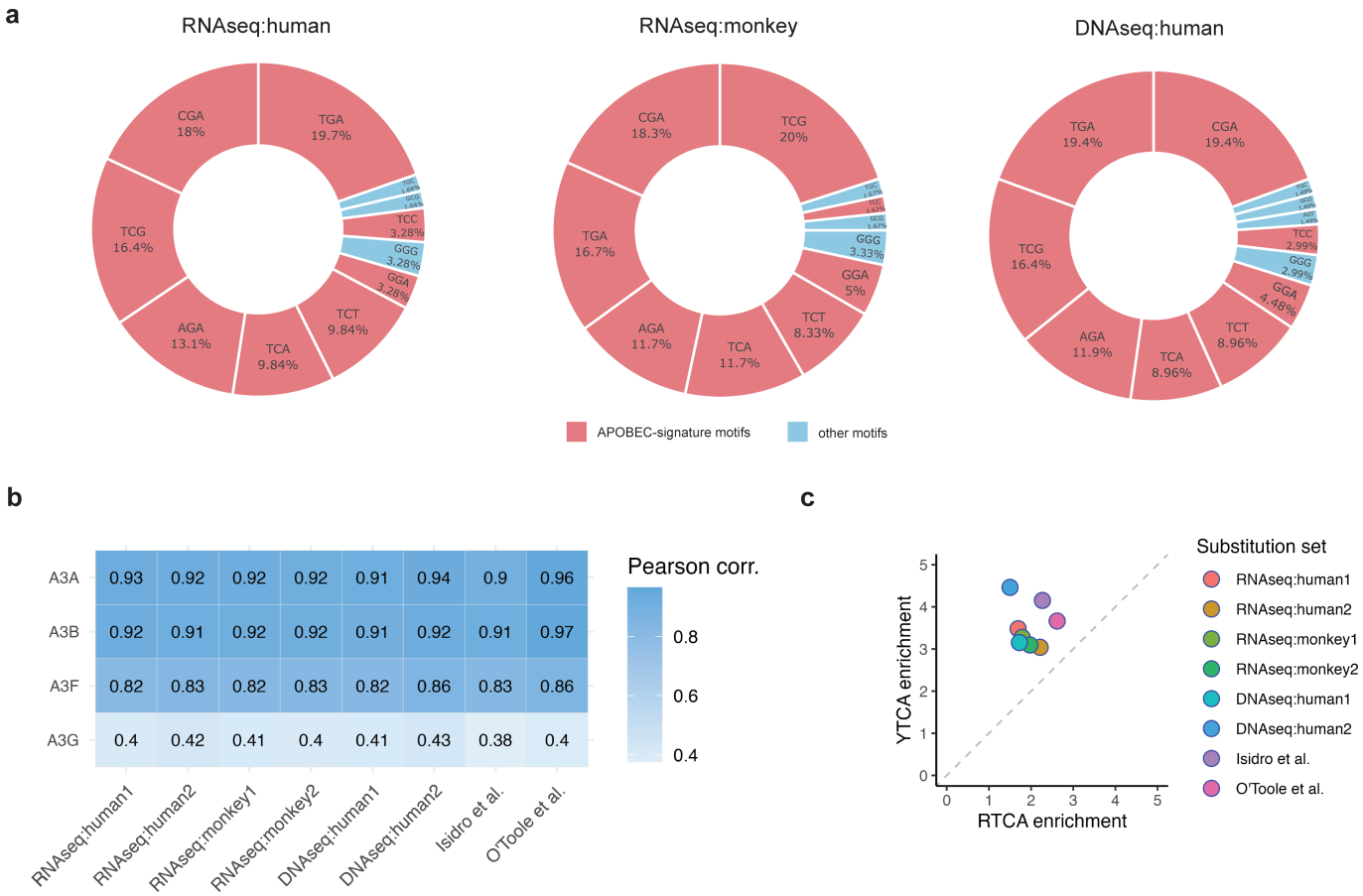

Figure 5. (**a**) *Analysis of the trinucleotide context of substitutions detected in RNA-seq and DNA-seq datasets from MPXV-infected human and monkey samples.* (**b**) *Correlation between the observed nucleotide context of substitutions and the known sequence preferences of APOBEC3 family enzymes. The highest correlation is observed with APOBEC3A (A3A) and APOBEC3B (A3B).* (**c**) *Discrimination between the A3A- and A3B-associated mutational profiles using YTCA vs. RTCA motif analysis, where YTCA is preferentially targeted by A3A and RTCA by A3B. All analyzed datasets showed a predominance of the YTCA motif, suggesting the involvement of A3A in the observed MPXV mutagenesis.*

from the 2022 outbreak [53,54] revealed that 52 positions (58%) overlapped with those found in the RNA-seq data (Fig. 4d). Notably, 38 of 41 positions (92%) reported in the earlier study by Isidro et al. [53] were also present in the union of sites detected in RNA-seq datasets and the O'Toole catalog [54].

To determine whether the observed substitutions occurred during the experiment or were already present, information on the specific MPXV strain used in each study was required. However, among the datasets considered, only the RNAseq:human2 project (PRJNA906618) provided information on the specific MPXV strain used. For this project, we excluded from the list of detected substitutions known mutations of this strain, resulting in the exclusion of 54 out of 77 substitutions. For other projects, we excluded mutations listed in the comprehensive catalog of mutations observed in MPXV strains [54]. Thus, we obtained a conservative estimate of mutations that occurred during the experiments conducted in the studies considered and were not present in the strains used (Supplemental file 3). Although the number of these mutations is relatively small, there is a clear enrichment of APOBEC-signature mutations (22 out of total 49, p=2.19×10-8). The distribution of these mutations across the projects is not uniform: three out of six projects contain most of the mutations, while the remaining three have very few or none.

Nucleotide context of substitutions suggests APOBEC3A as the likely enzyme responsible for APOBEC-signature mutagenesis

We further analyzed the trinucleotide context of APOBEC-signature substitutions identified in RNA-seq and DNA-seq studies (Fig. 5a). The most prevalent motif was TCG/CGA, accounting for 34% of all C→T/G→A substitutions in the human RNA-seq data, 38% in the non-human primates RNA-seq, and 35% in the DNA-seq datasets. This was followed by the TCA/TGA motif, which represented 30% of substitutions in the human RNA-seq, 28% in the non-human primates RNA-seq and 28% in DNA-seq samples. The TCT/AGA motif was the third most common motif, contributing 20% of substitutions in both human and non-human primates RNA-seq datasets and 21% in DNA-seq data. The fourth most frequent motif, TCC/GGA, accounted for only ~7% of substitutions – substantially lower than other TC-containing motifs. Notably, this distribution differs from patterns observed in APOBEC-related cancer studies [55], where mutations in cytosines in the TCA motif are typically the most prevalent.

We then analyzed the extended nucleotide context of APOBEC-signature substitutions to compare sequence profiles of these substitutions with known specificity models of APOBEC family members. First, we constructed positional weight matrices (PWMs) representing the sequence preferences of four APOBEC enzymes – A3A, A3B, A3G, and A3F – using available mutation data (see Methods). We then generated PWMs from the identified APOBEC-signature substitutions. The comparison of these matrices (Fig. 5b) revealed that the observed substitution profile most closely resembles the specificities of A3A and A3B. To differentiate between A3A and A3B, we conducted a YTCA vs. RTCA motif enrichment analysis [43], as these motifs are well-established signatures distinguishing these two enzymes. This analysis, applied across all datasets, consistently showed an enrichment of the YTCA motif (Fig. 5c), implicating A3A as the likely source of the observed mutations. These results align with previous observations for MPXV mutations [28], but contradict an experimental study [56] suggesting APOBEC3F as the enzyme responsible for APOBEC-signature mutations in MPXV.

Discussion

Interaction between viruses and host proteins play a crucial role in the viral life cycle. Viruses typically do not encode all proteins required for their life cycle, relying instead on host cellular proteins to perform essential functions. In contrast to interactions with recruited host proteins, viruses also engage in a constant battle with cellular defense factors — proteins whose primary function is to restrict viral replication. When a virus circulates within a host species for an extended period, it typically evolves mechanisms to evade cellular defenses. However, zoonotic transmission to a new host is usually associated with intensive interactions between the virus and the host's antiviral proteins. This was partially observed during the SARS-CoV-2 pandemic and has become more pronounced in the recent monkeypox outbreak. Specifically, an accelerated accumulation of mutations exhibiting the signature of the APOBEC3 subfamily of viral restriction factors was observed in monkeypox virus strains circulating in humans during the 2022 outbreak [53,54]. Given that APOBEC3 enzymes can target both ssDNA and RNA [21], and that MPXV genomes and transcripts are localized in the cytoplasm during the viral life cycle [13], it remains an open question whether MPXV transcripts are also subject to APOBEC-mediated mutagenesis. In this study, we analyzed publicly available MPXV RNA-seq data to address this issue.

We considered RNA-seq data from two human and two non-human primate studies and developed a pipeline for detecting potential RNA editing events. Our analysis yielded consistent findings across all considered datasets, revealing a moderate enrichment of APOBEC-signature substitutions in MPXV transcripts. However, the majority of these substitutions occurred at high frequencies at specific positions within transcripts, leading us to conclude that either they are fixed mutations at the DNA level, or that they are RNA editing events occurring recurrently at specific hotspots. To distinguish between these two possibilities, we performed a detailed analysis of the identified APOBEC-signature substitutions. Firstly, as APOBEC3 catalyzes C→T editing in mRNA, we found that many G→A substitutions, which are complementary to C→T changes, remained even after normalization relative to the gene strand orientation. Secondly, we found that the observed substitutions were more frequently located at synonymous positions than at non-synonymous ones. Thirdly, we analyzed the distribution of these substitutions along the MPXV genome and found no correlation with the locations of genes expressed at different stages of the transcriptional program (early, intermediate, or late). Fourthly, secondary structure patterns previously associated with APOBEC mutational hotspots were not observed at the substitution sites. Finally, we compared the sets of identified APOBEC-signature substitutions with known mutation sites in MPXV strains and found a substantial overlap. This overlap may, in part, reflect the presence of pre-existing mutations in the analyzed samples, as the specific viral strains used in the RNA-seq studies were not clearly specified. Taken together, these observations led us to conclude that the identified APOBEC-signature positions represent substitutions that occurred at the DNA level rather than RNA editing events.

We also investigated which member of the APOBEC family is most likely responsible for the observed mutagenesis. Given that members of the APOBEC family differ to some extent in their substrate specificity, we modelled the specificity via position-weight matrices using existing data for the A3A, A3B, A3F, and A3G enzymes. By comparing these models to the specificity profile derived from the observed data, we found consistently across all datasets that the sequence context of the substitutions best matches the specificity of A3A and A3B enzymes. Distinguishing between A3A and A3B specificity had been extensively considered in the context of cancer data [43], and here we applied a similar approach, identifying A3A as the likely mutator. The known subcellular localization of A3A and A3B supports this observation, as A3B is predominantly localized in the nucleus, whereas A3A has been detected in both the nucleus and cytoplasm [57].

Although members of the APOBEC family are known to restrict a wide range of viruses, including both positive- and negative-sense single-strand RNA viruses, two distinct mechanisms may be involved: a deaminase-dependent pathway, which leads to C-to-T hypermutation, and a deaminase-independent pathway, in which APOBEC proteins inhibit viral replication through direct binding [58]. The deaminase-independent mechanism has been experimentally confirmed for both DNA and RNA viruses. Although it is more frequently associated with DNA viruses and retroviruses, evidence for its occurrence in RNA viruses also exists [59,60], including reports of APOBEC-signature mutations in the positive-sense RNA virus SARS-CoV-2 [61,62]. Since APOBEC enzymes are known to mutate both viral DNA and RNA, the mechanisms preventing hypermutation of viral RNA transcripts remain poorly understood and require further research.

Acknowledgements

This study was supported by RSF grant #22-14-00132 (to D.N.I) and Assignment FFRW-2024-0004 (to G.V.P., E.V.M., and R.K.A.). We thank Irina Ponomareva for designing the preprint layout.

References

1. Lum FM, Torres-Ruesta A, Tay MZ, Lin RTP, Lye DC, Rénia L, et al. Monkeypox: disease epidemiology, host immunity and clinical interventions. Nature Reviews Immunology. Nature Research; 2022. pp. 597–613. doi:10.1038/s41577-022-00775-4
2. Magnus P von, Andersen EK, Petersen KB, Birch-Andersen A. A POX-LIKE DISEASE IN CYNOMOLGUS MONKEYS. Acta Pathologica Microbiologica Scandinavica. 1959; 46: 156–176. doi:https://doi.org/10.1111/j.1699-0463.1959.tb00328.x
3. Ladnyj ID, Ziegler P, Kima E. A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. Bull World Health Organ. 1972; 46: 593–597
4. Monzón S, Varona S, Negredo A, Vidal-Freire S, Patiño-Galindo JA, Ferressini-Gerpe N, et al. Monkeypox virus genomic accordion strategies. Nat Commun. 2024; 15. doi:10.1038/s41467-024-46949-7
5. Huang Y, Mu L, Wang W. Monkeypox: epidemiology, pathogenesis, treatment and prevention. Signal Transduction and Targeted Therapy. Springer Nature; 2022. doi:10.1038/s41392-022-01215-4
6. McFadden G. Poxvirus tropism. Nature Reviews Microbiology. 2005. pp. 201–213. doi:10.1038/nrmicro1099.
7. Karagoz A, Tombuloglu H, Alsaeed M, Tombuloglu G, AlRubaish AA, Mahmoud A, et al. Monkeypox (mpox) virus: Classification, origin, transmission, genome organization, antiviral drugs, and molecular diagnosis. Journal of Infection and Public Health. Elsevier Ltd; 2023. pp. 531–541. doi:10.1016/j.jiph.2023.02.003
8. Beer EM, Rao VB. A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. PLoS Negl Trop Dis. 2019;13: e0007791-. Available: https://doi.org/10.1371/journal.pntd.0007791
9. Bunge EM, Hoet B, Chen L, Lienert F, Weidenthaler H, Baer LR, et al. The changing epidemiology of human monkeypox—A potential threat? A systematic review. PLoS Negl Trop Dis. 2022;16: e0010141-. Available: https://doi.org/10.1371/journal.pntd.0010141
10. Otieno JR, Ruis C, Onoja AB, Kuppalli K, Hoxha A, Nitsche A, et al. Global genomic surveillance of monkeypox virus. Nat Med. 2024. doi:10.1038/s41591-024-03370-3
11. Nagy GÁ, Tombácz D, Prazsák I, Csabai Z, Dörmő Á, Gulyás G, et al. Exploring the transcriptomic profile of human monkeypox virus via CAGE and native RNA sequencing approaches. mSphere. 2024. doi:10.1128/msphere.00356-24
12. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat Med. 2022;28: 1569–1572. doi:10.1038/s41591-022-01907-y
13. Alakunle E, Kolawole D, Diaz-Cánova D, Alele F, Adegboye O, Moens U, et al. A comprehensive review of monkeypox virus and mpox characteristics. Frontiers in Cellular and Infection Microbiology. Frontiers Media SA; 2024. doi:10.3389/fcimb.2024.1360586
14. Hatmal MM, Al-Hatamleh MAI, Olaimat AN, Ahmad S, Hasan H, Ahmad Suhaimi NA, et al. Comprehensive literature review of monkeypox. Emerging Microbes and Infections. Taylor and Francis Ltd.; 2022. pp. 2600–2631. doi:10.1080/22221751.2022.2132882
15. Brennan G, Stoian AMM, Yu H, Rahman MJ, Banerjee S, Stroup JN, et al. Molecular Mechanisms of Poxvirus Evolution. mBio. American Society for Microbiology; 2023. doi:10.1128/mbio.01526-22
16. Shchelkunov SN, Totmenin A V., Safronov PF, Mikheev M V., Gutorov V V., Ryazankina OI, et al. Analysis of the monkeypox virus genome. Virology. 2002;297: 172–194. doi:10.1006/viro.2002.1446
17. Rubins KH, Hensley LE, Bell GW, Wang C, Lefkowitz EJ, Brown PO, et al. Comparative analysis of viral gene expression programs during poxvirus infection: A transcriptional map of the vaccinia and monkeypox genomes. PLoS One. 2008;3: 1–12. doi:10.1371/journal.pone.0002628
18. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol. 2010;27: 2038–2051. doi:10.1093/molbev/msq088
19. Gigante CM, Korber B, Seabolt MH, Wilkins K, Davidson W, Rao AK, et al. Multiple lineages of monkeypox virus detected in the United States, 2021-2022. Available: https://www.science.org
20. O'Toole Á, Neher RA, Ndodo N, Borges V, Gannon B, Gomes JP, et al. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. Science (1979). 2023;382: 595–600.
21. Salter JD, Bennett RP, Smith HC. The APOBEC Protein Family: United by Structure, Divergent in Function. Trends in Biochemical Sciences. Elsevier Ltd; 2016. pp. 578–594. doi:10.1016/j.tibs.2016.05.001
22. Mertz TM, Collins CD, Dennis M, Coxon M, Roberts SA. APOBEC-Induced Mutagenesis in Cancer. 2022. doi:10.1146/annurev-genet-072920
23. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: Footprints and mechanisms. Nature Reviews Cancer. Nature Publishing Group; 2014. pp. 786–800. doi:10.1038/nrc3816
24. Ratcliff J, Simmonds P. Potential APOBEC-mediated RNA editing of the genomes of SARS-CoV-2 and other coronaviruses and its impact on their longer term evolution. Virology. 2021;556: 62–72. doi:10.1016/j.virol.2020.12.018
25. Saini N, Gordenin DA. Hypermutation in single-stranded DNA. DNA Repair (Amst). 2020;91–92. doi:10.1016/j.dnarep.2020.102868
26. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science (1979). 2019;364. doi:10.1126/science.aaw2872
27. Langenbucher A, Bowen D, Sakhtemani R, Bournique E, Wise JF, Zou L, et al. An extended APOBEC3A mutation signature in cancer. Nat Commun. 2021;12. doi:10.1038/s41467-021-21891-0
28. Dobrovolná M, Brázda V, Warner EF, Bidula S. Inverted repeats in the monkeypox virus genome are hot spots for mutation. J Med Virol. 2023;95. doi:10.1002/jmv.28322
29. Leinonen R, Sugawara H, Shumway M. The sequence read archive. Nucleic Acids Res. 2011;39. doi:10.1093/nar/gkq1019
30. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44: D733–D745. doi:10.1093/nar/gkv1189
31. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191
32. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski

C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635
33. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26: 589–595. doi:10.1093/bioinformatics/btp698
34. Pedersen BS, Quinlan AR. Mosdepth: Quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34: 867–868. doi:10.1093/bioinformatics/btx699
35. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10. doi:10.1093/gigascience/giab008
36. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. NAR Genom Bioinform. 2021;3. doi:10.1093/nargab/lqab019
37. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. 2011. Available: http://www.tbi.univie.ac.at/RNA.
38. Kazanov FM, Matveev E V., Ponomarev G V., Ivankov DN, Kazanov MD. Analysis of the abundance and diversity of RNA secondary structure elements in RNA viruses using the RNAsselem Python package. Sci Rep. 2024;14. doi:10.1038/s41598-024-80240-5
39. Brázda V, Kolomazník J, Lýsek J, Hároníková L, Coufal J, Šťastný J. Palindrome analyser – A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Biochem Biophys Res Commun. 2016;478: 1739–1745. doi:10.1016/j.bbrc.2016.09.015
40. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho S-J, Malim MH. Cytidine Deamination of Retroviral DNA by Diverse APOBEC Proteins. Current Biology. 2004;14: 1392–1396. doi:10.1016/j
41. Wiegand HL, Doehle BP, Bogerd HP, Cullen BR. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO Journal. 2004;23: 2451–2458. doi:10.1038/sj.emboj.7600246
42. Langlois MA, Beale RCL, Conticello SG, Neuberger MS. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. Nucleic Acids Res. 2005;33: 1913–1923. doi:10.1093/nar/gki343
43. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47: 1067–1072. doi:10.1038/ng.3378
44. Huang Y, Bergant V, Grass V, Emslander Q, Hamad MS, Hubel P, et al. Multi-omics characterization of the monkeypox virus infection. Nat Commun. 2024;15. doi:10.1038/s41467-024-51074-6
45. Watanabe Y, Kimura I, Hashimoto R, Sakamoto A, Yasuhara N, Yamamoto T, et al. Virological characterization of the 2022 outbreak-causing monkeypox virus using human keratinocytes and colon organoids. J Med Virol. 2023;95. doi:10.1002/jmv.28827
46. Kakuk B, Dörmő Á, Csabai Z, Kemenesi G, Holoubek J, Růžek D, et al. In-depth Temporal Transcriptome Profiling of Monkeypox and Host Cells using Nanopore Sequencing. Sci Data. 2023;10. doi:10.1038/s41597-023-02149-4
47. Ye Q, Zhang D, Zhang RR, Xu Q, Huang XY, Huang B, et al. A penta-component mpox mRNA vaccine induces protective immunity in nonhuman primates. Nature Communications . 2024;15. doi:10.1038/s41467-024-54909-4
48. Langenbucher A, Bowen D, Sakhtemani R, Bournique E, Wise JF, Zou L, et al. An extended APOBEC3A mutation signature in cancer. Nat Commun. 2021;12. doi:10.1038/s41467-021-21891-0
49. Novoa RR, Calderita G, Arranz R, Fontana J, Granzow H, Risco C. Virus factories: associations of cell organelles for viral replication and morphogenesis. Biol Cell. 2005;97: 147–172. doi:10.1042/bc20040058
50. Deng Y, Navarro-Forero S, Yang Z. Temporal expression classes and functions of vaccinia virus and mpox (monkeypox) virus genes. mBio. 2025; e0380924. doi:10.1128/mbio.03809-24
51. Dobrovolná M, Brázda V, Warner EF, Bidula S. Inverted repeats in the monkeypox virus genome are hot spots for mutation. J Med Virol. 2023;95. doi:10.1002/jmv.28322
52. Martínez-Sanz J, Ponce-Alonso M, Martínez-García L, de la Cueva V, Olavarrieta L, Montero L, et al. Description of mpox reinfection by whole genome sequencing. International Journal of Infectious Diseases. 2023;137: 111–113. doi:10.1016/j.ijid.2023.10.017
53. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat Med. 2022;28: 1569–1572. doi:10.1038/s41591-022-01907-y
54. O'Toole A, Neher RA, Ndodo N, Borges V, Gannon B, Gomes JP, et al. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. Science (1979). 2023;382: 595–600. doi:10.1126/science.adg8116
55. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149: 979–993. doi:10.1016/j.cell.2012.04.024
56. Suspène R, Raymond KA, Boutin L, Guillier S, Lemoine F, Ferraris O, et al. APOBEC3F Is a Mutational Driver of the Human Monkeypox Virus Identified in the 2022 Outbreak. Journal of Infectious Diseases. 2023;228: 1421–1429. doi:10.1093/infdis/jiad165
57. Yang B, Li X, Lei L, Chen J. APOBEC: From mutator to editor. Journal of Genetics and Genomics. Institute of Genetics and Developmental Biology; 2017. pp. 423–437. doi:10.1016/j.jgg.2017.04.009
58. Zhu T, Niu G, Zhang Y, Chen M, Li CY, Hao L, et al. Host-mediated RNA editing in viruses. Biology Direct. BioMed Central Ltd; 2023. doi:10.1186/s13062-023-00366-w
59. Klimczak LJ, Randall TA, Saini N, Li JL, Gordenin DA. Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. PLoS One. 2020;15. doi:10.1371/journal.pone.0237689
60. Kockler ZW, Gordenin DA. From rna world to sars-cov-2: The edited story of RNA viral evolution. Cells. MDPI; 2021. doi:10.3390/cells10061557
61. Di Giorgio S, Martignano F, Gabriella Torcia M, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Sci Adv. 2020. Available: https://www.science.org
62. Wei L. Retrospect of the Two-Year Debate: What Fuels the Evolution of SARS-CoV-2: RNA Editing or Replication Error? Current Microbiology. Springer; 2023. doi:10.1007/s00284-023-03279-z