G3, 2026, 16(6), jkag080
https://doi.org/10.1093/g3journal/jkag080
Advance Access Publication Date: 28 March 2026
Investigation

APOBEC3A is the predominant global editor of cytosines in human mRNAs and in single-strand RNA viruses

Zachary W. Kockler,[1] Hamed Bostan,[2] Leszek J. Klimczak,[2] Yun-Chung Hsiao,[1] Matthew S. Dennen,[1] Molly E. Cook,[3] Tony M. Mertz,[4] Ludmila Perelygina ,[5] Marat D. Kazanov,[6] Jian-Liang Li ,[2] Steven A. Roberts,[4] Dmitry A. Gordenin [1,*]

[1]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, Durham, NC 27709, United States
[2]Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, Durham, NC 27709, United States
[3]Epigenetics and RNA Biology Laboratory, National Institute of Environmental Health Sciences, Durham, NC 27709, United States
[4]Department of Microbiology and Molecular Genetics, University of Vermont Cancer Center, University of Vermont, Burlington, VT 05405, United States
[5]Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, United States
[6]Faculty of Engineering and Natural Sciences, Sabanci University, Tuzla/İstanbul 34956, Türkiye

*Corresponding author: Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Durham, NC 27709, United States. Email: gordenin@niehs.nih.gov

APOBEC cytidine deaminases can convert cytosines to uracils in DNA as well as in RNA. The knowledge of DNA deamination motifs preferred by individual APOBECs revealed APOBEC3A as a major source of hypermutation in cancer. However, the extent and relative contribution of specific APOBECs into RNA editing remains unclear as their preferred RNA editing motifs have not been defined. Here, using a parallel DNA and RNA sequencing strategy, coupled with motif-centered statistical analyses, we sought to identify mRNA edits and diagnostic editing motifs in yeast and human cells overexpressing individual APOBEC enzymes. This approach revealed a prevailing global enrichment for the uCg trinucleotide motif with even greater preference to the motif's cytosines located in 3′ base of a loop within a hairpin-loop secondary structure when APOBEC3A, but not any other tested APOBEC, was overexpressed. Further analysis revealed the APOBEC3A-like diagnostic motif enrichment in editing calls from human cancers and blood cells. The APOBEC3A-like editing motif also prevailed in the RNA genomes of SARS-CoV-2 virus pandemic isolates, as well as in infectious persistent rubella viruses, and in polioviruses emerging from live-attenuated vaccine strains. Together, our results indicate that APOBEC3A is the predominant global APOBEC RNA editor with a potential to impact cell physiology and viral evolution.

Keywords: RNA editing; APOBEC; RNA viruses; transcriptome; cytidine deaminases; viral evolution; vaccine-derived viruses; single-strand DNA; viral genomes

Introduction

The apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 1 (APOBEC1) cytidine deaminase was discovered as a physiologically important programmed site-specific mRNA editor (Powell et al. 1987; Teng et al. 1993). Later studies revealed that other related enzymes encoded in the human APOBEC3 gene cluster can also perform untargeted conversion of cytosines into uracils in single-strand RNA as well as in single-stranded DNA (Refsland and Harris 2013; Lerner et al. 2018; Pecori et al. 2022 and references therein). In viral RNA genomes, the newly created uracil is maintained and inherited after replication reviewed in Kockler and Gordenin (2021). However, for DNA, the results of APOBEC cytidine deamination allow multiple outcomes. If the uracil remains until DNA replication, an adenine base will be placed across it, whereupon further replication will fix the uracil position as a thymine. Additionally, uracil DNA glycosylases can remove uracil bases in DNA to form an abasic site (AP-site), which can ultimately result in any of the 4 nucleotides filling that position (see Fig. 1 in Dennen et al. 2024 and references therein). Altogether, the resulting effects of APOBEC-mediated deamination can lead to DNA hypermutation or RNA hyperediting, which can result in high mutation load in cancers or in antiviral activity (Roberts et al. 2013; Green and Weitzman 2019; Alexandrov et al. 2020; Kockler and Gordenin 2021).

The roles of different APOBEC family members in DNA mutagenesis was established based on experimentally obtained knowledge of their preferences for oligonucleotide mutational motifs (3 to 4 nucleotides in size) centered around a mutated cytosine (Mertz et al. 2022). The preferred DNA mutational motif of most of the APOBEC family enzymes is the tCn (where n stands for any base and capitalized C is the deaminated cytosine). The preference to this motif was also clear in 2 single-base substitution mutational signatures (SBS2 and SBS13) agnostically extracted by nonnegative matrix factorization from large number of mutational catalogs of human tumors (Alexandrov et al. 2013, 2020). An exception, APOBEC3G has a preferred cCn trinucleotide motif (Conticello 2012; Refsland and Harris 2013; Harris and Dudley 2015; Green and Weitzman 2019). The choices of APOBECs potentially responsible for mutating cancer genomes were further narrowed down based upon APOBEC mRNA expression levels in tumor samples where APOBEC3B is the most expressed, while APOBEC3A mRNA correlates best with SBS2/13 loads, suggesting

Received on 27 February 2026; accepted on 23 March 2026
Published by Oxford University Press on behalf of The Genetics Society of America 2026.

that either or both are responsible for the cancer mutations (Burns, Lackey, et al. 2013; Burns, Temiz, et al. 2013). To better determine which APOBEC is responsible for the mutations in cancer, further refinement of the APOBEC mutational motif was performed in yeast ssDNA hypermutation model where either APOBEC3A or APOBEC3B was overexpressed. Preference to mutational motifs was statistically distinguishable in a −2 position upstream of the deaminated cytosine. Specifically, APOBEC3A preferred the ytCa motif (y = pyrimidine) while APOBEC3B had a preference to the rtCa motif (r = purine). Refined motif preference enabled to classify cancer samples with detectable APOBEC mutagenesis as APOBEC3A-like or APOBEC3B-like. It turned out that the APOBEC3A-like tumors had 10-fold more mutations than the APOBEC3B-like tumors (Chan et al. 2015). Together, this suggested that APOBEC3A is the prevailing DNA mutator in human cancers when compared with the other APOBECs. Later, this conclusion was confirmed in direct studies of cultured tumor cells and human tumors (Petljak et al. 2019, 2022; Dananberg et al. 2024). Thus, defining motif preference in a defined yeast model can enable identification of APOBEC enzyme(s) prevailing in cytidine deamination in real-life biological contexts.

Initially, APOBEC3 enzymes were considered to be exclusively DNA mutators (Refsland and Harris 2013). However, later studies have reported massive editing of cytosines in mRNAs in human cell cultures, as well as in human tumors. Further, editing was increased in response to hypoxia and interferons, in a number of genes important for cell function and for antiviral defense (Sharma et al. 2015, 2017, 2019). Massive RNA editing was detected when APOBEC3A was overexpressed (Sharma et al. 2017) or in tumors with increased APOBEC3A mutagenesis (Jalili et al. 2020) indicating capability of this enzyme to serve as a global RNA editor. Another candidate for global mRNA editing was a double-domain (dd) APOBEC3G, showing the same as in DNA cCn motif. APOBEC3G-associated edits in cells propagated in atmospheric oxygen concentration were limited to a smaller number of sites; however, edits became more widespread in hypoxic conditions (Sharma et al. 2016, 2019). Both APOBEC3A and APOBEC3G showed preference for loops in stem-loop RNA secondary structures, and APOBEC3A had even greater preference for cytosines on the 3′-border of the loop (Sharma and Baysal 2017; Jalili et al. 2020). Studies of Chen and colleagues utilizing experiments with reporter system or agnostic signature extraction from published data suggested high RNA-editing potential of APOBEC3A compared with other APOBECs (Kim et al. 2023; Fixman et al. 2024). There is a single report on potential mRNA editing by overexpressed APOBEC3B (Alonso de la Vega et al. 2023). Since this gene is overexpressed in many tumors as well as in normal tissues, it should also be included in the list considered as possible global RNA editors in vivo. While there are no data suggesting RNA editing by other APOBEC3 enzymes, their role cannot be excluded a priori. In summary, prior to our current study, it was clear that all or some of tC-specific APOBEC enzymes may account for C to U changes (editing) in human mRNAs; however, relative contribution in global mRNA editing was not clear.

APOBEC enzymes were also implicated in the editing of viral RNA genomes. We found earlier that most of the base substitutions in hypermutated vaccine-derived rubella virus isolates from granulomas of individuals with primary immunodeficiency (PID) were C to U in plus (genomic) strand. Up to 30% of all base substitutions were C to U changes in the uCn trinucleotide motifs conforming with a tCn motif preferred by most of APOBECs in DNA, while the cCn motif preferred by APOBEC3G was depleted (Perelygina et al. 2019). Further, RNA editing in infectious SARS-CoV-2 single-strand RNA suggested APOBEC's role prompted by uCn as the primary editing motif for cytosines (Di Giorgio et al. 2020; Klimczak et al. 2020; Azgari et al. 2021; Ratcliff and Simmonds 2021; Kim et al. 2022; Yang et al. 2025). In model experiments, transient transfection of cultured human cells containing SARS-CoV-2 with an APOBEC3A expression vector caused greater levels of C→U mutations in uCn motif compared with other APOBEC vectors (Nakata et al. 2023). However, the identity of the specific APOBEC(s) responsible for the bulk of mutagenic editing in genomes of RNA viruses infecting humans has not been revealed.

In both cases, mRNA editing and mutagenesis in the genomes of RNA viruses, more precise knowledge of specific oligonucleotide motifs preferred by different APOBEC enzymes can clarify their relative roles in editing. Here, we utilized a parallel RNA and DNA sequencing approach to reliably identify induced mRNA edits within yeast and human cells that hyperexpressed functional human APOBECs. These mRNA edits were then coupled with knowledge-based motif-centered statistical analyses to quantitatively define the preference of editing motif of each APOBEC. This strategy revealed that though each tested APOBEC can mutate DNA as well as edit RNA, APOBEC3A is the strongest RNA editor. We used the APOBEC3A-like RNA editing motif to assess its prevalence and enrichment within C→U RNA edits of cellular and viral RNAs. For that purpose, we explored published datasets of mRNA editing in human cancers and blood cells as well as in RNA virus genome editing for vaccine-derived rubella viruses (VDRVs), polioviruses, and in SARS-CoV-2. We found that in all the biological systems that we analyzed, there was an APOBEC3A-like motif preference, altogether, suggesting that APOBEC3A is not only the prevailing DNA mutator in human cancers but also the predominant global RNA editor among multiple APOBEC enzymes.

Materials and methods

Strains and plasmids

The yeast strains were isogenic to ySR128 (Roberts et al. 2012). Strains were MATalfa mating type carried *his7-2 leu2-3,112 trp1-289* mutations: *CAN1, ADE2,* and *URA3* were deleted from their original positions. In this background, the *lys2::ADE2-URA3-CAN1* array was inserted at the native *LYS2* on the right arm of chromosome II, ∼230 kb from the centromere and 340 kb from the telomere. An additional replacement deletion *ung1::NAT* eliminated uracil DNA glycosylase, thereby preventing creation of AP-sites at positions of cytosine deamination in DNA. One additional strain of the same lineage was created by introducing the replication protein A (RPA) DNA-binding-deficient *rfa1-t33* mutation, which was confirmed by Sanger sequencing. These 2 strains were transformed with a Hyg^R-marked autonomously replicating vector with a doxycycline-inducible full-size APOBEC inserted ORFs into the empty vector (EV) backbone: pSR4199 (EV), pSR433 (APOBEC1), pZAK031 (APOBEC3A), pSR440 (APOBEC3B), pSR469 (APOBEC3C), and pZAK045 (full-size dd, ddAPOBEC3G) (Mertz et al. 2025). Yeast strains and plasmids are listed in Supplementary Table 10. Autonomous replication status of vectors in transformants was confirmed by frequent loss of Hyg^R plasmid marker after propagation on the media without antibiotics. All *RFA1* wild-type yeast strains for experiments described in this paper were grown at 30 °C, and *rfa1-t33* strains were propagated at 23 °C except during short-term induction of APOBEC ORFs performed at 30 °C.

Parallel DNA and RNA sample collection in yeast

Maximizing the power of our parallel DNA and RNA sequencing approach required the collection of 2 yeast samples taken from the same culture in parallel. The yeast strains overexpressing APOBECs were streaked for single colonies on Hyg media directly from −80 °C stocks and grown for 2 d at 30 °C (*RFA1* wild-type yeast) or for 3 d at 23 °C (*rfa1*-t33 yeast). The entire single colony (~$1 \times 10^{e+07}$ cells) was suspended in 50 mL of liquid Yeast Extract Peptone Dextrose Adenine media supplemented with hygromycin 0.2 mg/mL to a density of $2 \times 10^{e+05}$ cell/mL and grown at 30 °C or at 23 °C (*RFA1-WT* or *rfa1-t33*, respectively) with agitation at 250 rpm to a cell density of ~$1.0 \times 10^{e+07}$ cells/mL (~12 to 14 h). For each APOBEC as well as for EV control, 16 to 24 cultures were grown. Once cultures reached designated density, doxycycline was added to final concentration of 10 µg/mL to stimulate APOBEC gene expression. Cells of all strains were then grown at 30 °C for 4 h (to allow 1 to 2 final cell division[s]) to stimulate expression of the APOBECs. Then, as quickly as possible, 2 mL of cells were aliquoted for RNA extraction, while 10 mL were aliquoted for DNA extraction. These aliquots were spun down, and the liquid was removed by aspiration leaving only the yeast cell pellet. The pellet was then flash frozen in liquid nitrogen and stored in −80 °C until extraction (typically within a week). The RNA extraction was performed using the Direct-zol RNA MiniPrep from Zymo Research, Irvine, CA, USA (Cat # R2052). DNA extraction was performed using the YeaStar Genomic DNA Kit from Zymo Research (Cat # D2002) following the standard chloroform version of the protocol, with the modification of doubling the digestion buffer and lysis buffer volumes. In addition to DNA extraction from bulk cultures, genomic DNA was also extracted from single colonies of canavanine-resistant mutants (16 for EV, 24 for APOBEC3C, and 36 for ddAPOBEC3G).

Parallel DNA and RNA sample collection in human cells

Cell culture and creation of stable cell populations

BT474 cells were obtained directly from ATCC as part of a breast cancer cell line panel, ATCC 30-4500K. BT474 and derivative cell populations were propagated according to ATCC guidelines in Hybri-Care media with 1.5 g $NaHCO_3$ per liter and 10% FBS. Stably transduced BT474 cells were selected with 20 µg/mL blasticidin, 500 µg/mL hygromycin B, or 1 µg/mL puromycin. HEK293T and HEK293T-TETR cells (described in Cortez et al. 2019) were grown in Dulbecco's Modified Eagle Medium (DMEM) + 10% Fetal Bovine Serum (FBS) or DMEM + 10% FBS with 5 µg/mL blasticidin, respectively. List of human cell lines can be found in Supplementary Table 10.

BT474 derived cell populations were established by serial lentiviral transductions as follows. Lentivirus to establish TETR expression was produced by cotransfection of pLenti_CMV_TetR_Blast (Addgene, Watertown, MA, USA, Cat #17492) into HEK293T cells with psPAX2 (Addgene, Cat#12260) and pMD2.G (Addgene, Cat #12259). Lentiviruses for pTM406 (EV), pTM664 (APOBEC3A overexpression), and pTM666 (APOBEC3B overexpression) were produced similarly but used HEK293T-TETR cells, in which APOBEC expression is repressed by TETR, which allows more efficient lentivirus production by limiting APOBEC-mediated restriction. To generate BT474 cells with doxycycline-inducible EV control, APOBEC3A or APOBEC3B, cells were first transduced with pLenti_CMV_TetR_Blast to establish TETR expression. Post selection, these cells were subsequently transduced with either pTM406, pTM664, or pTM666 to introduce the EV control, APOBEC3A, or APOBEC3B coding sequences. For transductions, lentiviral-containing supernatant was added to BT474 cells in the presence of Lentiblast Premium (OZ Biosciences, San Diego, CA, USA), which was diluted 1:200 in the cell culture media. Following transductions, the media was changed at 16 h, and selection started at 72 h. The plasmids pTM664 and pTM666 were previously described and confirmed to produce active protein by purification and in vitro deaminase assay (Cortez et al. 2019). pTM406 was constructed from an LR Clonase (Thermo Fisher, Waltham, MA, USA, Cat# 11791020) reaction between pLenti PGK Puro DEST (Addgene, Cat #19068) and pENTR1A no ccDB (w48-1) (Addgene, Cat #17398) and sequence validated by Sanger sequencing.

Human cell line preparation for next-generation sequencing

Expression of APOBEC3A and APOBEC3B in BT474 + pLenti_CMV_TetR_Blast, pTM664 and BT474 + pLenti_CMV_TetR_Blast, pTM666 cells were induced by addition of 1.0 µg/mL doxycycline to the cell culture media 96 h before harvesting cells. BT474 + pLenti_CMV_TetR_Blast, pTM406 cells were similarly treated and used as EV controls. RNA and DNA were harvested from asynchronous cell populations at ~60% confluence from two 10 cm dishes for each preparation. RNA was purified using the E.N.Z.A. Total RNA Kit (Omega Bio-Tek, Norcross, GA, USA, Cat# R6834-00) followed by DNase I treatment, phenol chloroform extraction, and ethanol precipitation. DNA was purified using the DNA Blood and Tissue kit (QIAGEN, Germantown, MD, USA). Isolated DNA was subsequently purified from a contaminating nuclease by phenol chloroform extraction and ethanol precipitation. The integrity and purity of isolated RNA and DNA was confirmed by visualization on an agarose gel with 1% vol/vol bleach or standard agarose gel stained with GelRed, respectively. Induction of APOBEC3A and APOBEC3B by doxycycline treatment was confirmed by qRT-PCR using oligo pairs oTM-746:oTM-748 (for APOBEC3A), oTM-737:oTM-740 (for APOBEC3B), and oTM-197:oTM-207 (for TBP gene) (Supplementary Fig. 2) as described in Cortez et al. (2019).

DNA mutation calling in yeast

DNA mutations were called after sequencing at the National Institute of Environmental Health Sciences (NIEHS) epigenomics core facility where DNA libraries were prepared using the Illumina DNA Prep kit and sequenced on the NextSeq or NovaSeq platforms. High mutagenic activity of APOBEC3A and APOBEC3B in yeast even without induction of Tet-promoter allowed to detect sufficient number of APOBEC-induced mutations from whole-genome sequencing of DNA prepared from bulk cultures. Mutagenesis with APOBEC1, APOBEC3C, and ddAPOBEC3G was weaker (Dennen et al. 2024; Supplementary Fig. 1); therefore, we used sensitized background *rfa1-t33* and prepared DNA from single-cell canavanine-resistant colonies isolated from the same bulk cultures that were used for high-coverage DNA and mRNA sequencing. The Illumina sequencing read files were imported into CLC Genomics Workbench (Version 20.0, QIAGEN) for read mapping to the ySR128 yeast reference genome (GenBank accessions CP036470.1 to CP036486.1 and Roberts et al. 2012) with duplicate reads being removed. Mutation calls for DNA motif analyses from whole-genome sequencing were obtained using CLC Genomics Workbench. For bulk cultures expressing highly mutagenic APOBEC3A and APOBEC3B, calls were obtained using basic Variant Caller option with subsequent filtering to a minimum allelic frequency of 20% in order to limit DNA mutation motif analysis to APOBEC-induced mutations occurring during the

first 1 to 2 cell divisions of the cell at the origin of the culture. Multistep workflow is detailed in Supplementary Table 11. Mutation calls from single-cell canavanine-resistant colonies from APOBEC1, APOBEC3C, and ddAPOBEC3G culture were obtained using multistep workflow centered around the Fixed Ploidy Variant Detection tool (ploidy of 1) with subsequent filtering to a minimum allelic frequency of 80%. The multistep workflow can be found in Supplementary Table 12. Distributions of variant allelic frequencies (VAFs) for single-base substitutions can be found in Supplementary Table 13. For both workflows, the resulting mutations were then filtered to remove calls that occurred in multiple samples to limit analysis to independent events. These filtered mutation call datasets were then listed in mutation annotation format (MAF). General description of MAF format is in https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format. MAF files were then processed using the P-MACD pipeline (https://github.com/NIEHS/P-MACD). All MAF files fed into P-MACD and their derivative versions *.anz4, containing +/−20 nucleotide flanks around each base substitution, are listed in Supplementary Table 2 and can be found at https://doi.org/10.5281/zenodo.18079216.

Parallel DNA and RNA sequence reads processing for RNA edit calling

Reads from paired DNA- and RNA-seq samples (submitted to SRA: Bio-project PRJNA1398811 for human samples and Bio-project PRJNA1398896 for yeast samples) were evaluated for quality and subjected to trimming to remove adapters and low-quality base sequences using TrimGalore (Krueger 2023). High-quality cleaned DNA-seq reads were aligned to their respective reference genomes (ySR128 [Roberts et al. 2012] for yeast samples and human hg19 for human samples) using minimap2 (Li 2018) (version 2.24-r1122) with the parameters –MD -ax sr. Similarly, high-quality cleaned RNA-seq reads were aligned to their corresponding reference genomes using HISAT2 (Zhang et al. 2021) (version 2.2.1) with the parameters –dta.

Aligned reads from both DNA-seq and RNA-seq data (in SAM format) were processed using the Picard toolkit (Way 2022) (version 2.26.11). The following steps were performed:

1) AddOrReplaceReadGroups: Appropriate read groups were added, and SAM files were converted to sorted BAM files with the parameters: VALIDATION_STRINGENCY = SILENT, SO = coordinate, and RGPL = illumina.
2) MarkDuplicates: PCR duplicates were marked using the parameters: CREATE_INDEX = true and VALIDATION_STRINGENCY = SILENT.

The resulting paired samples (aligned, sorted, indexed, and duplicate-marked DNA and RNA BAM files) were subsequently used as input for RNA-editing analysis with the REDITools (Picardi 2024) pipeline.

RNA editing calling

RNA edits for both yeast and human were identified using REDItoolDnaRna.py from the REDITools (Picardi 2024) (version V1) pipeline with the parameters: -s 2 -t 20 -n 0.05 -N 0.05 -e -E -d -D. For human genome, hg19 reference was used. The reference annotation ySR128_v2.gff file (https://doi.org/10.5281/zenodo.18079122) for yeast ySR128 genome was used for accurate transcript mapping. DNA positions used to identify RNA editing calls were filtered to positions where only cytosines were present in DNA reads to limit potential DNA mutations being called as RNA edits. RNA edits were only called at positions containing only cytosines in DNA reads, while cytosines or thymines were called in those positions. This resulted in the production of the final high-confidence C to U RNA editing call set.

MAF files generation for RNA edits

The final filtered sets of RNA editing calls for yeast and human experiments for each sample were converted into the MAF format for the downstream analysis similar to that outlined for DNA mutation calls with additional information relevant for RNA edit filtering included (Supplementary Table 2). The RNA editing calls in these MAF files were then filtered to include only editing calls where more than one read had evidence of a cytosine to uracil (displayed as thymine) RNA edit (abbreviated to rTg1). Yeast calls were further filtered to remove EV calls from calls in APOBEC overexpressing cells to account for potential calls coming from a non-APOBEC cytosine deamination source. Human cell experiments were analyzed without further filtering because EV cells contain all APOBECs as well as non-APOBEC sources of potential deamination. Even so, EV cells had a low enrichment in motifs characteristic of APOBEC3A or APOBEC3B (see Results). MAF files for mRNA editing events from publications (Sharma et al. 2015; Perelygina et al. 2019; Jalili et al. 2020; Klimczak et al. 2020) were generated directly from reported edit calls. MAF files for RNA edits in vaccine-derived isolates of ss(+)RNA viruses, Rubella (Wanat et al. 2022), or poliovirus (Cherkasova et al. 2002; Kew et al. 2002, 2004; Yang et al. 2005, 2010) and unpublished GenBank entries (Supplementary Table 9) were generated by comparing GenBank genome sequences of a virus isolate with a GenBank reference of the parent virus, using the procedure described in Perelygina et al. (2019) and Klimczak et al. (2020).

Datasets of APOBEC-induced mutations in ssDNA

APOBEC-induced mutations formed in yeast subtelomeric ssDNA

We utilized a dataset of mutations induced by either APOBEC3A or APOBEC3B expressed in long stretches of subtelomeric ssDNA formed by 5′ to 3′ resection of *ung1Δ* yeast carrying *cdc13-1* defect in telomere capping (Fig. 2a; and Chan et al. 2015). Mutation catalogs were taken directly from (Chan et al. 2015) and reorganized as *anz4 version of MAF format (Supplementary Table 2). Similar to Chan et al. (2015), Saini et al. (2020), and Hudson et al. (2023) APOBEC-induced mutations in C:G pairs formed at the distance <30 kb from the left telomeres were assigned to cytosines of the ssDNA in the unresected bottom strand, while mutations at the distance <30 kb from the right telomeres were assigned to cytosines of the ssDNA in the unresected top strand. MAF files with strand-assigned APOBEC-induced mutations were then used to calculate enrichments with all possible 192 mutational motifs as described in (Klimczak et al. 2020; Hudson et al. 2023).

C- of G-strand-coordinated mutation clusters formed in long persistent stretches of ssDNA in human cancers

We utilized the output of APOBEC mutagenesis analysis from Data Resources of Pan-cancer Analysis of Whole Genomes (PCAWG) (see Supplementary Table 4 in Consortium 2020). Previously, we have demonstrated that several types of human cancers included in PCAWG dataset—Bladder (Bladder-TCC), Breast (Breast-AdenoCA and Breast-LobularCA), Cervical (Cervix-SCC), Head and Neck (Head-SCC), Lung (Lung-AdenoCA and Lung-SCC) (PCAWG abbreviated names shown in parentheses)—are highly enriched in APOBEC mutational motif and can be

even separated into APOBEC3A-like and APOBEC3B-like tumors by detailed motif analysis (Roberts et al. 2013; Chan et al. 2015). Moreover, we reveled that in APOBEC hypermutated tumors, C- or G-strand-coordinated clusters containing more than 3 mutations always demonstrate higher enrichment of APOBEC mutational motif than the scattered mutations in C:G pairs. All other C:G-preferring motifs, eg the nCg to nTg motif of meCpG mutagenesis were depleted in C- or G-strand-coordinated clusters (Roberts et al. 2013; Roberts and Gordenin 2014; Chan and Gordenin 2015; Kazanov et al. 2015; Gerhauser et al. 2018). Therefore, we filtered mutations in APOBEC-hypermutated types of PCAWG cancers to only C:G pair mutations in C- or G-strand-coordinated clusters containing more than 3 mutations of tumors that can also be assigned to either APOBEC3A-like or to APOBEC3B-like categories. We used *anz4 MAF files from these filtered mutation catalogs to calculate enrichments with all possible 192 trinucleotide mutational motifs similar to the analysis of APOBEC mutagenesis in long persistent subtelomeric ssDNA of yeast hyperexpressing either APOBEC3A or APOBEC3B (see previous section). Mutations in G-strand-coordinated clusters were assigned to cytosine deamination in complementary strand.

Knowledge-based statistical valuation of trinucleotide-centered enrichments for RNA edits and DNA mutations to derive signatures

The calculation of the overrepresentation of the mutational or editing motifs was performed using the patterns of mutagenesis by APOBEC cytidine deaminases (P-MACD) pipeline (previously described in Chan et al. (2015); Roberts et al. (2013); Saini et al. (2020); Hudson et al. (2023) and available at https://github.com/NIEHS/P-MACD). For independent cultures of the yeast strain carrying the same plasmid, DNA mutation or RNA editing calls were indicated separately in the mutation annotation file (MAF) but pulled into a single input MAF for P-MACD pipeline. This P-MACD pipeline calculates the enrichment of a specific trinucleotide motif when compared with random mutagenesis/editing in each MAF based upon the assumption that a specific motif would occur more frequently among mutated/edited bases than in the surrounding sequence context (context is defined by the +/− 20 bases surrounding the mutated/edited base). The equation to calculate enrichment of a given mutation signature is provided below for the DNA mutational motif tCn to tTn as an example (the mutated nucleotide is capitalized in the motif; n corresponds to any nucleotide).

$$\text{Enrichment (tCn} \rightarrow \text{tTn)} = \frac{[\text{Mutations}_{\text{tCn}\rightarrow\text{tTn}}] \times [\text{Context}_{\text{C}}]}{[\text{Mutations}_{\text{C}\rightarrow\text{T}}] \times [\text{Context}_{\text{tCn}}]}$$

To determine if there is a statistically significant increase in the enrichment for a tested trinucleotide motif, Fisher's exact test for odds ratios was performed wherein the ratio of the number of mutations within the trinucleotide motif ($\text{Mutations}_{\text{tCn}\rightarrow\text{tTn}}$) and those that do not conform to the trinucleotide motif ($[\text{Mutations}_{\text{C}\rightarrow\text{T}}] - [\text{Mutations}_{\text{tCn}\rightarrow\text{tTn}}]$) was compared to the number of bases in the context that were in the trinucleotide motif ($\text{Context}_{\text{tCn}}$) vs those that were not in the motif ($[\text{Context}_{\text{C}}] - [\text{Context}_{\text{tCn}}]$).

After MAF files were generated for APOBEC-induced ssDNA mutations C to T or RNA editing calls C to U, P-MACD was run for tCa motif common for APOBEC3A and APOBEC3B (Chan et al. 2015). Among the P-MACD generated files, there were *anz4, which contain all information of the input MAF file plus several additional columns including the column with the +/−20 bases surrounding the mutated/edited base. This column was essential for calculating enrichments as well as *P*-values for all possible trinucleotide mutational motifs. When a DNA strand containing nucleotide with a mutagenic lesion is known, enrichment in 192 trinucleotide motifs can be calculated. If a nucleotide change cannot be assigned to 1 of the 2 strands, 96 motif enrichments, each totaling enrichments for 2 reverse complements, are calculated. We calculated enrichments with all possible trinucleotide mutational or editing motifs in ssDNA or in ssRNA, respectively, using the procedure described in Klimczak et al. (2020) and Hudson et al. (2023). Since there are only 16 possible C to T or C to U of APOBEC-induced mutations or RNA edits, respectively, we applied correction of *P*-values for 16 independent hypotheses using the Benjamini–Hochberg method. For trinucleotide motifs where the enrichment was >1 and the corrected *P*-value was <0.05, odds ratios were compared by Breslow–Day test for homogeneity to evaluate statistical significance of difference between enrichments (Liu 2005).

RNA editing density in secondary structures

The density of RNA editing was calculated as the ratio of C-to-T substitutions within tCn motifs to the total number of potential APOBEC targets, eg the number of tCn motifs in the RNA molecule being analyzed:

$$\text{Density} = N_{\text{edits}}/N_{\text{trg}},$$

where N_{edits} represents the number of RNA edits, and N_{trg} represents the total number of tCn motifs in a target. Calculations for specific positions within a hairpin-loop structure were performed as described in Ponomarev et al. (2022). Human noncanonical DNA structures, including hairpins, were taken from Non-B DB database (Cer et al. 2013). For human genome, the calculation of targets was limited to exon regions associated with a transcript of gene where edit(s) were called. For genes producing several transcripts, the transcript with the maximum number of edited exons was selected; if multiple transcripts had the same number of edited exons, the transcript with the highest number and the total length of unedited exons was chosen.

Results

Experimental design for parallel detection of APOBEC-induced mutagenesis in DNA and mRNA editing in yeast

There are many contributing factors that make identifying mRNA edits more difficult than DNA mutations (reviewed in Kockler and Gordenin 2021). Firstly, each RNA transcript is transient within the cell, so a sequencing run will generate a "snapshot" of the current mRNA molecules and any change introduced by editing would disappear when the edited mRNA molecule ceases to exist. On the contrary, changes (mutations) in genomic DNA persist and replicate. Secondly, each editing change in mRNA would exist in a small fraction of transcripts within a cell, while genomic DNA mutations are present in 50% or even 100% allelic fraction in a diploid or in a haploid genome, respectively. Moreover, DNA mutations are transmitted to mRNA transcripts and thus obscure RNA edit calling. Altogether, detection of genome-wide (global) RNA editing requires highly accurate sequencing, large sequence coverage, and high efficiency of DNA mutation filtering. To facilitate accurate detection of mRNA edits induced by functional human

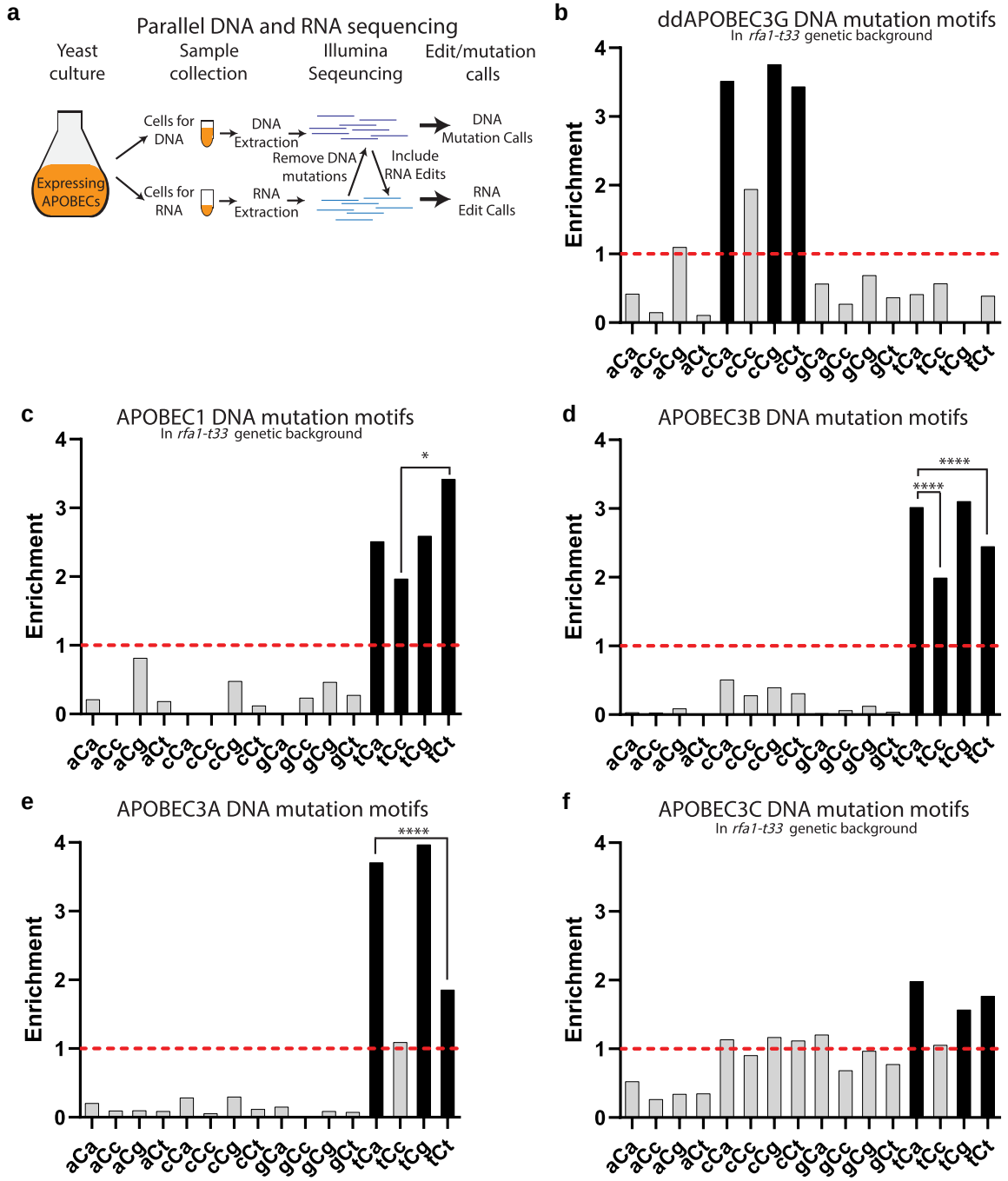

Fig. 1. Motif preferred by DNA mutations induced by human APOBECs expressed in yeast. a) Schematic of the parallel DNA and mRNA sequencing strategy for identifying mRNA editing events. Yeast cells are collected from the same culture and processed for DNA and RNA extraction, followed by the whole-genome or whole-transcriptome sequencing. To prevent DNA mutations from being called as RNA edits, positions of the DNA genome with at least one read with a mutation were not considered. RNA sequencing reads were aligned to annotated gene transcripts, and RNA edits occurring at expected cytosines were identified (see Materials and Methods). b to f) DNA reads from bulk cultures expressing stronger APOBEC3A and APOBEC3B (*RFA1-WT* background) or from *can1* mutants isolated from cultures expressing weaker APOBEC1, APOBEC3C, and ddAPOBEC3G (*rfa1-t33* background) served to call genome-wide mutations, which were used to identify preferred motifs of DNA mutagenesis by each APOBEC. Enrichments of the 16 possible trinucleotide DNA mutational motifs centered around C to T (including reverse complements) are shown on each panel. Enrichment values shown on each panel can be found in Supplementary Table 3. Dashed lines indicate enrichment level of 1.0. Black bars indicating a statistically significant enrichment ($q \leq 0.05$ determined for 16 *P*-values by Benjamini–Hochberg); gray bars—not significant. Enrichments for all possible trinucleotide mutational motifs, including *P*-values and *q*-values can be found in Supplementary Table 4. Pairs of odds ratios of all motif–motif combinations within a 4-motif uCn group were compared by Breslow–Day test for homogeneity of the odds ratios (Supplementary Table 5). Pairs with statistically significant differences are highlighted by brackets. Level of significance is shown above brackets (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$). b) ddAPOBEC3G, c) APOBEC1, d) APOBEC3B, e) APOBEC3A, f) APOBEC3C.

APOBECs in the budding yeast, *Saccharomyces cerevisiae*, we utilized a strategy of a parallel DNA and RNA sequencing (Lerner et al. 2021) (Fig. 1a; Materials and Methods). This strategy included overexpressing functional APOBECs in our yeast strains for just 1 to 2 cell generations followed by parallel DNA and RNA extraction from the same culture (Fig. 1a) and Illumina

sequencing at high coverage (target coverage at least 200× for DNA as well as for RNA). For DNA, the high coverage facilitated detection of de novo mutations that were present in small fractions of cells, while in RNA, it gives the greatest possibility to detect rare editing events. In order to reduce the chance of false calls, high-quality RNA edits were called only at positions where cytosines are expected in the RNA transcripts with at least 2 independent instances of a cytosine to uracil edit. Further, these RNA edits were then cross-referenced with DNA sequencing to remove any position in the transcriptome where there was at least one DNA read with a mutation in the same position as the suggested RNA edit. This limited the possibility of DNA mutations that are transcribed into mRNA and incorrectly called as RNA edits.

The first requirement for the use of an APOBEC ORF for detection of mRNA editing in yeast was the production of deaminase with the in vivo enzymatic activity. For that purpose, we relied on detection of DNA mutagenesis caused by expression of that ORF. In our prior study, we assessed the in vivo DNA mutagenesis in yeast by every human APOBEC with demonstrated in vitro deaminase activity (Dennen et al. 2024). We found statistically significant mutagenesis in the *CAN1* loss-of-function mutation reporter caused in wild-type yeast by full-size APOBEC3A, APOBEC3B ORFs and by partial single-domain APOBEC3G. We also documented mutagenesis caused by APOBEC1 and APOBEC3C in strains carrying *rfa1-t33*, a hypomorph mutation in RPA subunit, which does not bind to DNA as readily as a wild-type RPA. Increased mutagenesis in *rfa1-t33* background was likely due to RPA impeding the access for APOBECs to cytosines in ssDNA (Wong et al. 2021 and references therein). Our prior studies of APOBEC3G mutagenesis in yeast were performed with the partial single-domain APOBEC3G ORF (Chan et al. 2012, 2013); however, RNA editing activity in human cells has been reported for full-size dd APOBEC3G (Sharma et al. 2016). Thus, the full-size ORF was used in this study to assess mRNA editing motif preference. Since RPA does not bind efficiently to mRNA when compared with ssDNA (Mazina et al. 2020), we do not anticipate any effect of *RFA1* genotype on RNA editing by APOBECs. We found that mutagenesis by ddAPOBEC3G increased around 4-fold in *rfa1-t33* vs *RFA1-WT* yeast and was also 3-fold increased over EV control (Supplementary Fig 1 and Table 1). Therefore, parallel detection of DNA mutagenesis and mRNA editing for full-size ddAPOBEC3G ORF was also performed in *rfa1-t33* strains. Mutagenesis by APOBEC1, APOBEC3C, and ddAPOBEC3G was detectable in *rfa1-t33*; however, it was lower than by APOBEC3A or APOBEC3B. Thus, we sequenced DNA from bulk cultures expressing these weaker APOBECs for parallel analysis with RNA reads, but we also sequenced genomes of single colony Can^R isolates for determining spectra of DNA mutations (see Materials and Methods).

Motif preference by APOBEC-induced DNA mutations in yeast and in human cancers

Using our parallel analysis strategy (Fig. 1a; Materials and Methods), a range of 200 to 2,000 DNA mutations were called in each group of yeast cultures expressing APOBEC3A, APOBEC3B, ddAPOBEC3G (dd), APOBEC1, and APOBEC3C (the latter 3 in a sensitized *rfa1-t33* background) (Supplementary Table 2 for file description and https://doi.org/10.5281/zenodo.18079216 for complete lists of mutation calls). In addition to statistically significant mutagenesis in reporter, mutagenic activity of APOBECs was supported by increased fractions among all base substitutions of C→T mutations (55% to 98% when compared with 47% in EV), and even more so by high fractions of APOBEC-specific mutagenesis motifs, 49% to 89% of tCn for APOBEC1, APOBEC3A, B,C, and 39% for of cCn for ddAPOBEC3G when compared with 9% to 13% in EV (calculated values can be found in Supplementary Table 4). However, percentages of mutations are not the most statistically rigorous way to determine mutational motif preference because they do not account for nucleotide and motif composition of genomic segments in which mutation happened. In order to determine motif preferences of DNA mutagenesis caused by human APOBECs in replicating yeast cells, we used knowledge-based motif-centered statistical analyses (Materials and Methods) to calculate and statistically evaluate enrichments of all possible dsDNA 96 trinucleotide motifs (including reverse complements) or 192 trinucleotide ssDNA motifs in DNA mutation calls from each data cohort (Supplementary Tables 3 and 5). Since APOBECs cause mutations by C to U deamination in ssDNA as well as in RNA, we then concentrated here and all subsequent analyses of DNA mutagenesis and RNA editing on 16 trinucleotide motifs centered around C→T (or C to U) changes in strains expressing each APOBEC (Fig. 1b to f; Materials and Methods). The ddAPOBEC3G that had enriched DNA mutational motifs within the group of 4 cCn trinucleotides (Fig. 1b), which was consistent with previous studies of APOBEC3G mutagenesis in yeast (Chan et al. 2012; Taylor et al. 2013). Each of the other 4 APOBECs tested had enriched trinucleotide DNA mutational motifs only within the group of 4 tCn trinucleotides (Fig. 1c to f), which is in agreement with prior observations for DNA mutagenesis with these enzymes (summarized in Mertz et al. 2022). Enrichment with motifs characteristic for each individual enzyme indicated that each of these proteins, even the weakest APOBEC1 and APOBEC3C, had cytidine deaminase activity in cultures where RNA editing activity and motif specificity were aimed to be accessed (Fig. 1a).

There were significant differences for enrichments within 4-motif tCn group of APOBECs (Fig. 1c to f). Importantly, APOBEC3A and APOBEC3B preferences of tCa over tCt motif were in agreement with prior studies of APOBEC mutagenesis in yeast subtelomeric ssDNA formed by the multikilobase 5′ to 3′ resection as well as in APOBEC mutation clusters formed in transient stretches of ssDNA in cancer genomes (Chan et al. 2015; Sakofsky et al. 2019). In the above cited studies, APOBEC3A and APOBECB motif-centered analysis was limited to tCa and to the 2-motif group tCw (tCa and tCt; w = A or T). We did not include tCc and tCg in prior analyses, where mutagenesis in yeast ssDNA was compared with genome-wide mutagenesis in cancers, because the tCc was likely to overlap with cCn mutagenesis—a motif preferred by APOBEC3G and the tCg with cytosine deamination in meCpG mutagenesis (nCg motif). However, these concerns are not valid for motif analyses in yeast expressing either APOBEC3A or APOBEC3B, because meCpG mutagenesis or APOBEC3G mutagenesis would not be expected. Therefore, we extended analysis of yeast ssDNA mutation catalogs obtained in Chan et al. (2015) to all 16 motifs centered around C to T substitution (Fig. 2a to c). In that work, we also established preference of APOBEC3A to extended motif ytCa (y = C or T), while APOBEC3B has a preference to rtCa (r = A or G). This approach enabled to assign APOBEC hypermutated TCGA; ICGC and PCAWG tumors to APOBEC3A-like or to APOBEC3B-like categories (Chan and Gordenin 2015; Sakofsky et al. 2019; Consortium 2020). We have also demonstrated that in cancers, C- or G-strand-coordinated clusters containing more than 3 mutations are highly enriched with APOBEC mutational motifs (note: based on APOBEC motifs enrichment, G-strand-coordinated clusters have originated from cytosine deamination in ssDNA formed by complementary

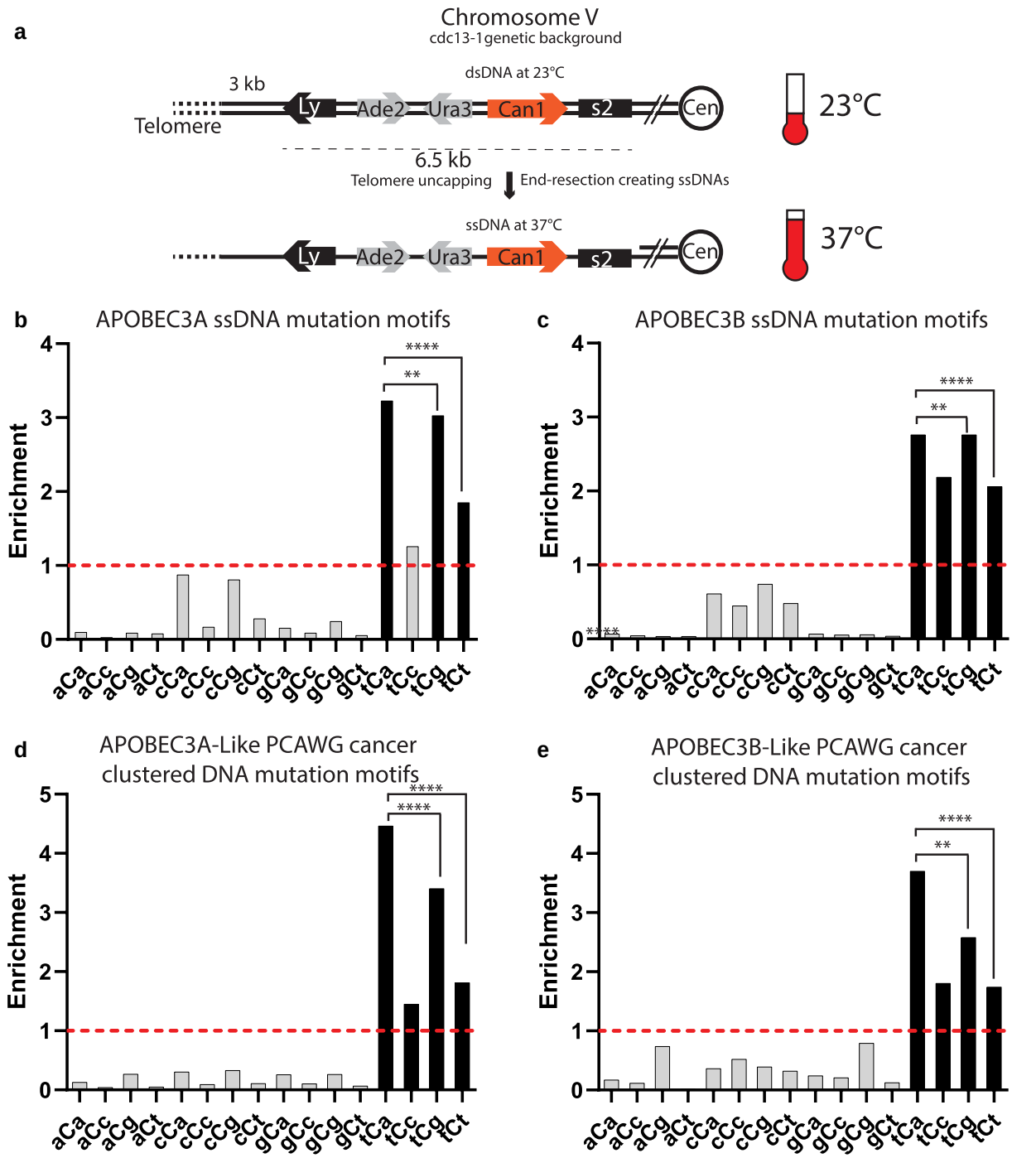

Fig. 2. Motif preference for APOBEC3A and APOBEC3B mutagenesis in ssDNA formed in yeast and in human cancers. a) Schematic of the subtelomeric triple reporter system in Chan et al. (2015) that exposes long stretches of ssDNA upon the introduction of the 37 °C nonpermissive temperature. b to e) Enrichment of the 16 possible C to T trinucleotide DNA mutational motifs and statistical evaluations are presented in the same format as in Fig. 1b to f. Enrichment values shown on each panel can be found in Supplementary Table 3. Dashed lines indicate enrichment level of 1.0. Black bars indicating a statistically significant enrichment ($q \leq 0.05$ determined for 16 *P*-values by Benjamini–Hochberg); gray bars—not significant. Enrichments for all possible trinucleotide mutational motifs, including *P*-values and *q*-values can be found in Supplementary Table 6. Pairs of odds ratios of all motif–motif combinations within a 4-motif uCn group were compared by Breslow–Day test for homogeneity of the odds ratios (Supplementary Table 5). b) Human APOBEC3A and c) human APOBEC3B expressed in the subtelomeric ssDNA system (data from Chan et al. 2015). Note: While tCg enrichment (2.76116) only slightly exceeded tCa enrichment (2.76062), statistically significant *P*-value ($P = 0.0048$, Supplementary Table 5) could be inflated by high counts of observed events. d) Enrichment of the 16 possible C to T DNA mutational motifs in Pan-Cancer Analysis of Whole Genomes tumors with an APOBEC3A-like whole-genome mutation motif preference and e) with an APOBEC3A-like mutation motif preference (data from Alexandrov et al. 2020; Consortium 2020).

strand). Thus, we used mutation catalogs of C- or G-coordinated long mutation clusters from APOBEC-hypermutated cancer types subdivided into APOBEC3A-like and APOBEC3B-like subgroups of tumors to evaluate enrichments with all 16 motifs (Fig. 2d and e). In both yeast subtelomeric ssDNA and in long ssDNA hypermutated in cancers, enrichment was observed only in motifs

belonging to tCn group, and tCa enrichments exceeded enrichments with tCt motif, which was similar to motif specificity in replicating yeast (compare Fig. 1d and e and Fig. 2b to e). However, unlike in replicating yeast, tCa enrichment statistically significantly exceeded tCg enrichment for APOBEC3A-induced ssDNA mutations in yeast and in cancers (Fig. 2b and d; Supplementary Table 6) and for APOBEC3B-induced ssDNA mutations in cancer mutation clusters (Fig. 2e). Differences with replicating yeast can be explained by several factors, including greater amounts or strength of RPA binding to short in size and in persistence ssDNA gaps formed during DNA replication when compared with long and persistent ssDNA formed in uncapped telomeres or at yet unknown points of tumor development (see Dennen et al. 2024 and therein). There was only slight, albeit statistically significant excess of tCg over tCa enrichment for APOBEC3B-induced mutations (2.76116 vs 2.76062, respectively) in yeast subtelomeric ssDNA system (Fig. 2c).

Motif preference of APOBEC induced mRNA editing in yeast

We next performed high-coverage mRNA sequencing from the same yeast cultures in which APOBEC mutagenesis in DNA was explored (Fig. 1a). Stringent filtering of calls was set to exclude any potential contamination with low-VAF mutation calls in DNA. For that purpose, mRNA C to U changes at positions in which there was at least a single DNA read containing the matching mutation were excluded. In addition, C to U edits were called as edits only if they were detected in at least 2 reads. We found 8,000 to 10,000 potential mRNA C to U edits in each set of cultures overexpressing APOBEC3A, APOBEC3B, ddAPOBEC3G, APOBEC1, and APOBEC3C (Supplementary Table 2; Materials and Methods). We then calculated enrichment for each of 16 trinucleotide motifs centered around C to U mRNA edits. The C to U mRNA changes in cultures expressing weaker editors—APOBEC1, APOBEC3C, and ddAPOBEC3G—showed small albeit statistically significant enrichments in 6 to 8 out of 16 motifs (Fig. 3a to c; Supplementary Table 7). Only one of these motifs in ddAPOBEC3G-expressing yeast fell into the cCn group exclusively enriched in ssDNA of ddAPOBEC3G cultures (compare Fig. 1b and Fig. 3a). The 2 tCn-specific enzymes, APOBEC1 and APOBEC3C, also did not show exclusive enrichment for the group of uCn motifs in RNA (see Fig. 1c vs Fig. 3b and Fig. 1f vs Fig. 3c). Thus, we cannot reliably assign C to U RNA changes to deamination by any of these weaker APOBECs.

Unlike the weaker APOBECs, C to U mRNA changes in yeast expressing APOBEC3A or APOBEC3B showed the same strong (for APOBEC3B) or exclusive (for APOBEC3A) preference to the group of uCn editing motifs selfsame with the tCn group of mutational motifs exclusively enriched in genomic DNAs of strains used for parallel detection of APOBEC DNA mutagenesis and RNA editing (compare Fig. 1d and e vs Fig. 3d). Thus, RNA-specific base substitution calls in APOBEC3A- and APOBEC3B-expressing yeast can be ascribed to RNA editing with greater confidence. Importantly, there were clear differences between APOBEC3A and APOBEC3B mRNA editing motif enrichments (Fig. 3d). Firstly, mRNA edits in APOBEC3A-expressing yeast showed a clear increase in uCa and uCg enrichments over mRNA edits by APOBEC3B. Secondly, uCg enrichment of RNA edits was statistically significantly greater than uCa enrichment in APOBEC3A but did not show statistically significant excess in APOBEC3B-overexpressing yeast. Importantly, statistically significant excess for preference to uCg over uCa in APOBEC3A yeast mRNA edits (Fig. 3d, left panel) was opposite to statistically significant preference of tCa over tCg in mutagenesis within persistent subtelomeric ssDNA in yeast and in APOBEC3A-induced mutation clusters in human cancers (Fig. 2b). The highly significant preference of tCa vs tCg was also found for APOBEC3A-like strand-coordinated DNA mutation clusters in APOBEC-hypermutated human cancers (Fig. 2d).

Establishing the uCg over uCa preference in C to U mRNA changes of APOBEC3A overexpressing yeast and the lack of such a preference in APOBEC3B-overexpressing yeast can aid in future analysis of mRNA edits in various biological contexts. Should this difference hold in APOBEC3A- and APOBEC3B-overexpressing human cells, this could aid in ascribing a specific APOBEC enzyme to RNA editing.

Motif and structural preferences in mRNA editing by APOBEC enzymes expressed in human cells

After establishing that human APOBEC3A and APOBEC3B not only caused DNA mutations but also can act as mRNA editors with distinguishable motif specificities, we explored mRNA editing by these APOBECs in human cells. For that purpose, we induced overexpression of human APOBEC3A (A3A-expression) or APOBEC3B (A3B-expression) in the BT474 breast cancer cell line, which accumulates ongoing mutation from endogenous APOBEC3A and APOBEC3B expression (Petljak et al. 2022). Doxycycline induction elevated APOBEC3A and APOBEC3B 500-fold and 7-fold, respectively, over endogenous transcript levels (Supplementary Fig. 2). This induction resulted in APOBEC3A expression being 0.33 that of APOBEC3B. While the spectrum and motif preference of DNA mutation calls were not analyzed, we subjected the doxycycline-induced APOBEC3 cell lines to a parallel DNA and RNA high-coverage sequencing strategy similar to the design with yeast cell cultures (Materials and Methods; Fig. 1a). Stringent filtering of mRNA changes resulted in sets of 3,000 to 4,000 C to U changes in mRNA of APOBEC-expressing cell lines (Supplementary Table 2; Materials and Methods). Motif analysis revealed that enriched motifs for both A3A and A3B primarily fell into uCn motifs (with 1 exception in A3B expressing cell line). There was a striking similarity with mRNA editing caused by these enzymes in yeast (compare Figs. 3d and 4a; Supplementary Table 7). Firstly, enrichment with uCn motifs was significantly higher in A3A vs A3B. Secondly, there was strong preference for uCg over uCa mRNA editing in A3A-expressing cell line.

It was previously established that human APOBEC3A has preference for 3′C in the loop of stem-loop structures formed by inverted repeats (IRs) in RNA as well as in DNA (Buisson et al. 2019; Jalili et al. 2020; Oh and Buisson 2022). We found similar preference for C to U mRNA edits in A3A but not in A3B cell line (Fig. 4b; Supplementary Table 8). Moreover, A3A edits in uCg had the highest densities in 3′C positions of IR loop when compared with other uCn motifs (Fig. 4c; Supplementary Table 8), which is also consistent with the suggestion that most of mRNA edit events in A3A-expressing cell line indeed were generated by A3A.

C to U transcriptome-wide mRNA editing motif preference in human cancers and in blood cells indicates prevailing role of APOBEC3A

The prevailing role of APOBEC3A in generating genome-wide mutation load in cancer genomes has been established in several studies using the role of germline genotype and distinct motif preference of this enzyme in cytosine deamination in DNA (Nik-Zainal et al. 2014; Chan et al. 2015; Petljak et al. 2022). Moreover, the in vitro preference of APOBEC3A to the loop in stem-loop structures is also detectable in APOBEC3A

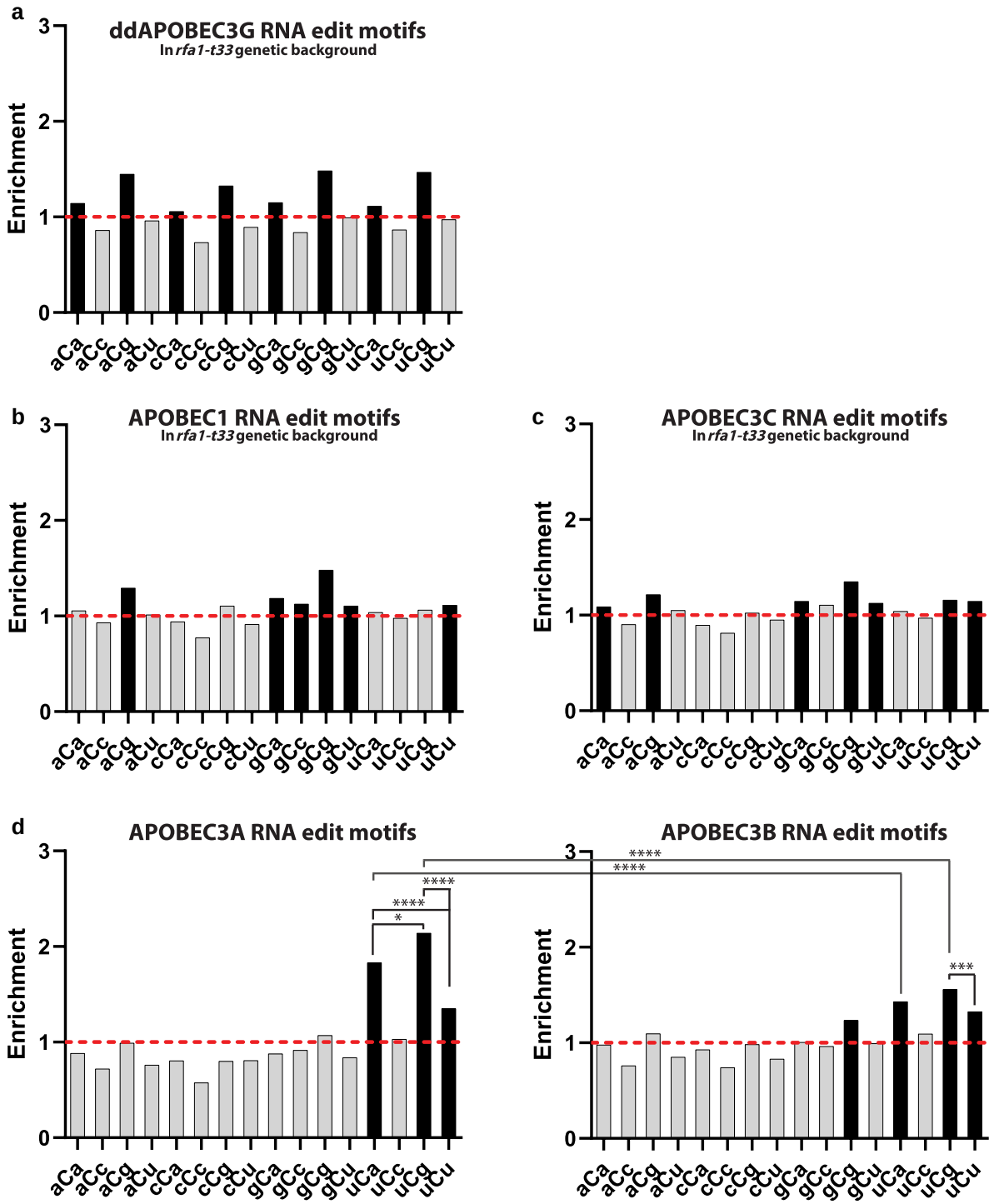

Fig. 3. Motif preference of APOBEC-induced mRNA edits in yeast. a to d) Enrichment of the 16 possible C to U trinucleotide mRNA editing motifs and statistical evaluations are presented in the same format as in Fig. 1b to f for DNA editing motifs. Enrichment values shown on each panel can be found in Supplementary Table 3. Dashed lines indicate enrichment level of 1.0. Black bars indicating a statistically significant enrichment ($q \leq 0.05$ determined for 16 *P*-values by Benjamini–Hochberg); gray bars—not significant. Enrichments for all possible trinucleotide mRNA editing motifs, including *P*-values and *q*-values can be found in Supplementary Table 7. Pairs of odds ratios of all motif–motif and cohort–cohort combinations (APOBEC3A vs APOBEC3B) within a 4-motif uCn group were compared by Breslow–Day test for homogeneity of the odds ratios (Supplementary Table 5). a) ddAPOBEC3G. b) APOBEC1. c) APOBEC3C. d) APOBEC3A (left panel) and APOBEC3B (right panel).

hypermutated cancers (Buisson et al. 2019; Ponomarev et al. 2022). When the preference for loops in RNA stem-loop structures has been explored in whole-exome sequenced tumors from TCGA cohort, such a preference was distinctly associated with APOBEC-hypermutated cancer samples (Jalili et al. 2020). We applied motif-centered analysis to over 59,000 transcriptome-wide RNA editing events in a previously reported mixed cohort of 25 tumors with high and 25 tumors with low levels of genome-wide APOBEC mutagenesis (Jalili et al. 2020). Indeed, enrichment analysis conformed with both features of APOBEC3A RNA editing established in our current study—there was exclusive enrichment in uCn group of motifs and the highest

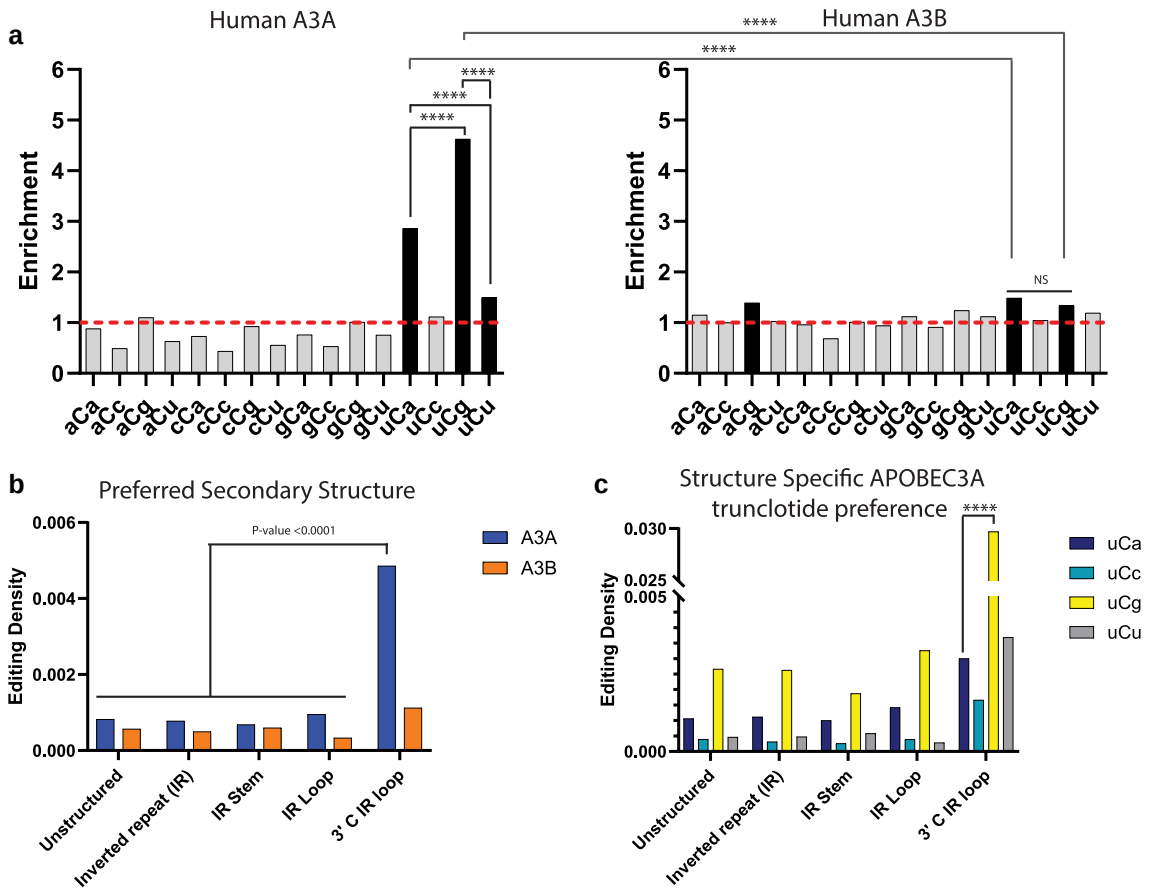

Fig. 4. Analysis of APOBEC-induced mRNA editing in human breast cancer cell line. a) Enrichment of the 16 possible C to U mRNA editing trinucleotide motifs in human BT-474 breast cancer cell lines expressing APOBEC3A (left panel) or APOBEC3B (right panel). mRNA editing motifs and statistical evaluations are presented in the same format as in Fig. 1b to f for DNA editing motifs. Enrichment values shown on each panel can be found in Supplementary Table 3. Dashed lines indicate enrichment level of 1.0. Black bars indicating a statistically significant enrichment ($q \leq 0.05$ determined for 16 *P*-values by Benjamini–Hochberg); gray bars—not significant. Enrichments for all possible trinucleotide mRNA editing motifs, including *P*-values and *q*-values can be found in Supplementary Table 7. Pairs of odds ratios of all motif–motif and cohort–cohort combinations (APOBEC3A vs APOBEC3B) within a 4-motif uCn group were compared by Breslow–Day test for homogeneity of the odds ratios (Supplementary Table 5). b) The density of APOBEC uCn mRNA editing motif in different secondary structure contexts in cell lines expressing APOBEC3A (A3A) and APOBEC3B (A3B). c) Densities of each of the 4 APOBEC3A mRNA editing motifs belonging to the 4-motif uCn group in each type of secondary structure. b and c) Brackets show significant differences between densities calculated by chi-square test. Source data for b) and c) and statistical comparisons can be found in Supplementary Table 8.

enrichment within this group was with uCg motif (Fig. 5a; Supplementary Table 7).

We next performed motif-centered analysis of the dataset of mRNA edits in noncancerous blood cells (Sharma et al. 2015). This study demonstrated transcriptome-wide C to U mRNA editing in hypoxia-treated monocytes and in monocyte-derived macrophages. Since editing in small number of hotspots was correlated with APOBEC3A expression, the authors proposed that this enzyme may also prevail in global mRNA editing. We applied motif-centered analysis to datasets of around 3,000 C to U editing events in monocytes and 140 edits in macrophages (Supplementary Table 2). In monocytes as well as in macrophages, uCg motif was enriched way more than any of the other 15 other C to U motifs (Fig. 5b and c; Supplementary Table 7). Another enriched motif in both datasets was uCa, also falling into uCn 4-motif group characteristic of APOBEC mutagenesis in DNA, which was also detected by our analysis in controlled APOBEC expression in yeast and in human cells (Figs. 3 and 4). C to U edits in monocytes also showed enrichment of cCg motif. Interestingly, the same motif was the only detectable outside the uCn group in macrophages. This could be due to either broader motif specificity of APOBEC3A or to the presence of the other low-capacity C to U editor. More studies are needed to confirm that cCg is indeed a new editing motif. Altogether, motif-centered analysis of global C to U editing calls in mRNAs of human cancers and in human blood cells revealed prevalence of APOBEC3A diagnostic editing features suggesting that this enzyme is a prevailing RNA editor in humans.

Enrichment with diagnostic uCg motif indicates a prevailing role for APOBEC3A in editing of (+)ssRNA viral genomes

RNA editing is not limited to mRNA and can occur in other RNA species. In RNA viruses that lack DNA intermediates in their life cycles, changes in viral RNA genomes stemming from various sources, including C to U edits by APOBEC enzymes, may affect viral fitness and/or evolution (Kockler and Gordenin 2021 and therein). RNA edits in viral genomes are propagated in progeny viruses and thus can be accurately identified by comparison with the genome of original viral strain providing the original genome sequence is known. The most straightforward detections and assignment of base editing to RNA stand can be done with (+)ssRNA viruses (outlined in Fig. 1 of Kockler and Gordenin 2021). However, for most human viruses with either RNA or

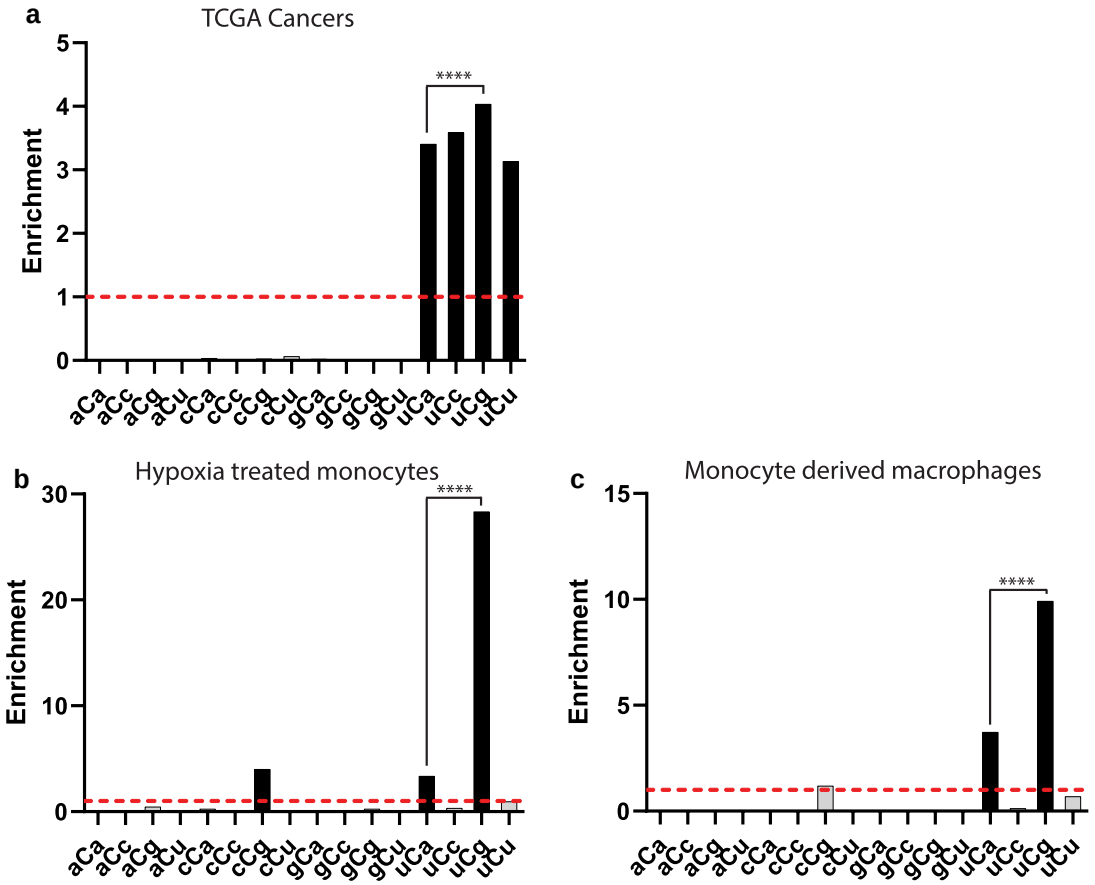

Fig. 5. Enrichments for trinucleotide mRNA editing motifs centered around C to U changes in TCGA cancers and in normal blood cells. a to c) Enrichment of the 16 possible C to U trinucleotide mRNA editing motifs and statistical evaluations are presented in the same format as in Fig. 1b to f for DNA editing motifs. Enrichment values shown on each panel can be found in Supplementary Table 3. Dashed lines indicate enrichment level of 1.0. Black bars indicating a statistically significant enrichment ($q \leq 0.05$ determined for 16 *P*-values by Benjamini–Hochberg); gray bars—not significant. Enrichments for all possible trinucleotide mRNA editing motifs, including *P*-values and q-values can be found in Supplementary Table 7. Pairs of odds ratios of all motif–motif within a 4-motif uCn group were compared by Breslow–Day test for homogeneity of the odds ratios (Supplementary Table 5). a) C to U editing motifs of mRNA from cancer genomes identified in Jalili et al. (2020). C to U editing motifs of mRNA from hypoxia-treated monocytes b) and from monocyte-derived macrophages c) (data from Sharma et al. 2015).

DNA genomes, identifying the progenitor viral genome by sequencing clinical isolates is challenging due to high genetic variability resulting in heterogeneous populations of viral quasispecies (Domingo et al. 2012; Domingo and Perales 2018, 2019).

Nevertheless, there are situations where a progenitor RNA viral genome is known. One such scenario occurs when a live-attenuated (weakened) vaccine virus persists in the host and undergoes RNA editing enabling within-the-host viral evolution. VDRVs were found in individuals with PID (Perelygina et al. 2020). These patients received the live-attenuated rubella vaccine strain RA27/3 prior to diagnosis of PID. Over time, some individuals developed skin granulomas which contained VDRVs with multiple, up to 300 de novo mutations within a ~10-kb (+)ssRNA genome. These mutations occurred between vaccination and collection of granuloma biospecimens. Motif analyses of VDRV base substitutions revealed that the predominant source of ssRNA genome editing were uCn-specific APOBEC cytidine deaminases (Perelygina et al. 2019). Here, we performed a motif-centered enrichment analysis with each possible trinucleotide motif centered around 586 C-to-U changes in (+)ssRNA VDRVs reported in that study. We found diagnostic features of APOBEC3A RNA editing, including predominant enrichment with uCn motifs and the highest enrichment in uCg motif (Fig. 6a; Supplementary Table 7). Moreover, 290 C-to-U edits in 3 RNA genomes from VDRVs recovered in another study from idiopathic skin granulomas of clinically immunocompetent individuals (Wanat et al. 2022) also exhibited clear diagnostic features of APOBEC3A editing (Fig. 6b; Supplementary Table 7). In addition, modest enrichment in aCg motif was observed in both immunocompromised and immunocompetent individuals (Fig. 6a and b; Supplementary Table 7). It remains unclear whether this is a minor motif of APOBEC3A-mediated editing or its enrichment is stemming from another, yet unknown editing mechanism.

The Sabin oral poliovirus vaccine (OPV) is another example of (+)ssRNA live-attenuated virus that can persist and acquire mutations in immunocompromised individuals resulting in emergence of vaccine-derived polioviruses (VDPVs). Some nucleotide substitutions in the VDPV genome resulted in reversion to a wild-type phenotype, leading to several poliomyelitis outbreaks in recent years (Kew et al. 2004; Devaux et al. 2023). We analyzed 59 GenBank entries of OPV VDPVs full genome sequences (Cherkasova et al. 2002; Kew et al. 2002, 2004; Yang et al. 2005, 2010; Liu et al. 2015) and unpublished GenBank entries (Supplementary Table 9; Materials and Methods), which altogether contained 6,253 C-to-U edits in (+)ssRNA genomes (Fig. 6c; Supplementary Table 7). Similar to rubella VDRV, uCg enrichment exceeded the other 3 motifs of uCn 4-motif group. Interestingly, another similarity with rubella virus was

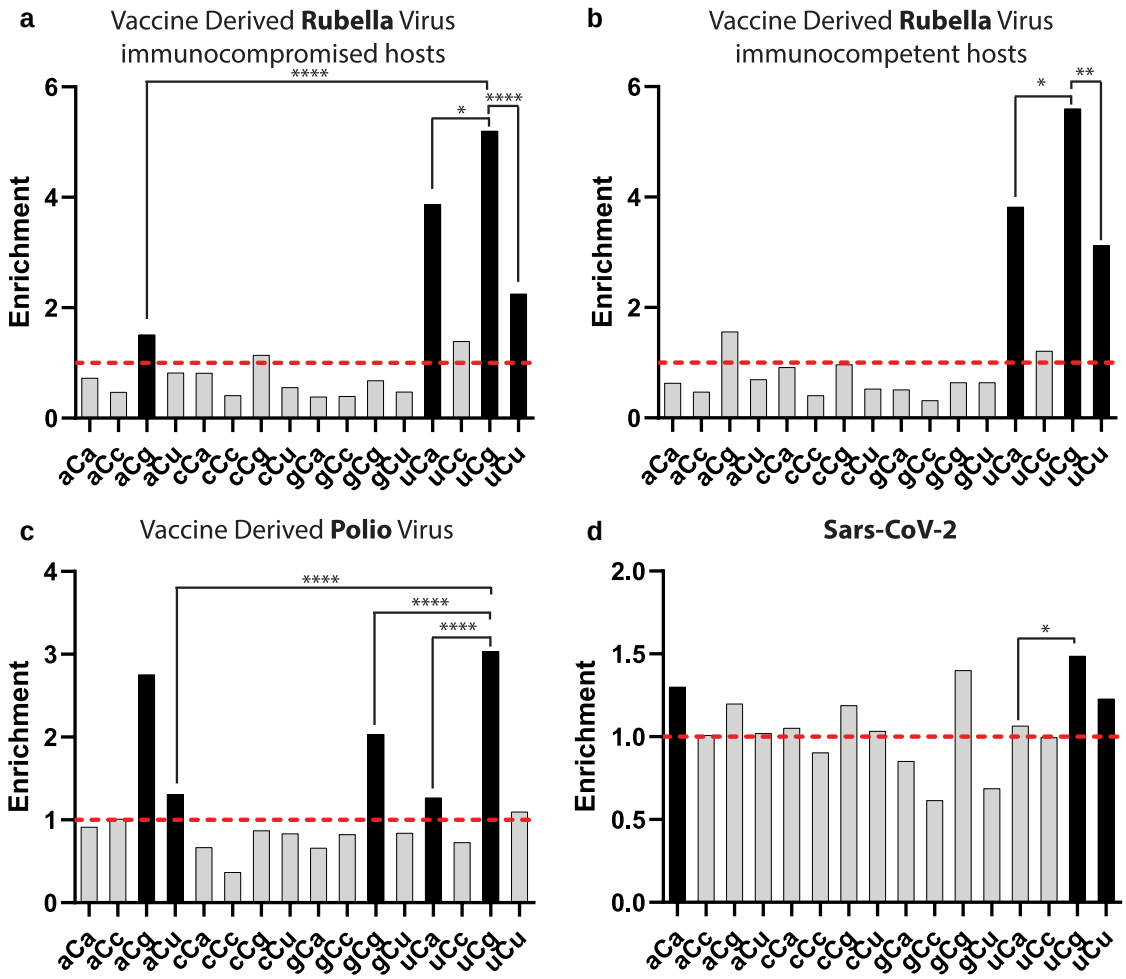

Fig. 6. Enrichments for trinucleotide mRNA editing motifs centered around C to U changes in genomes of (+)ssRNA viruses. a to d) Enrichment of the 16 possible C to U trinucleotide mRNA editing motifs and statistical evaluations are presented in the same format as in Fig. 1b to f for DNA editing motifs. Enrichment values shown on each panel can be found in Supplementary Table 3. Dashed lines indicate enrichment level of 1.0. Black bars indicating a statistically significant enrichment ($q \leq 0.05$ determined for 16 *P*-values by Benjamini–Hochberg); gray bars—not significant. Enrichments for all possible trinucleotide mRNA editing motifs, including *P*-values and *q*-values can be found in Supplementary Table 7. Pairs of odds ratios of all motif–motif within a 4-motif uCn group and with statistically significant enrichments outside the uCn group were compared by Breslow–Day test for homogeneity of the odds ratios (Supplementary Table 5). a) Vaccine-derived rubella from immunocompromised individuals (data from Perelygina et al. 2019). b) Vaccine-derived rubella from immunocompetent individuals (data from Wanat et al. 2022). c) Vaccine-derived polio virus. Accession numbers and publication references can be found in Supplementary Table 9. d). Genomes of SARS-CoV-2 pandemic isolates (data from Klimczak et al. 2020).

statistically significant enrichment with aCg motif not belonging to APOBEC-specific uCn group. Enrichment with aCg was comparable with uCg enrichment (Supplementary Table 5; Breslow–Day *P*-value of difference between 2 odds ratios = 0.167). Statistically significant enrichments with 2 other non-uCn motifs, aCu and gCg were much lower than uCg enrichment (Fig. 5c). Altogether, motif-centered analyses of nucleotide substitutions in 2 vaccine-derived viruses, poliovirus, and rubella virus indicate the prevailing role of APOBEC3A-like editing motif preference with potential involvement of other editor(s).

Another biological context where editing of (+)ssRNA viral genomes in multiple isolates can be traced to a single progenitor strain is represented by sequenced SARS-CoV-2 isolates collected during the COVID-19 pandemic. Here, we analyzed RNA editing events that had the highest likelihood to escape selection pressure, which alter the frequency of variants in pandemic population and thus make the spectrum of editing events recovered from human population to differ significantly from the spectrum of initial editing. For that purpose, we included only C to U edits in nonfunctional parts of the virus as well as synonymous changes and also counted each individual edit only once. The latter also helped to eliminate confounding effects of edits that have occurred early and propagated in human population throughout the course of pandemic. Such filtering produced 1,375 C to U edits (Klimczak et al. 2020; Supplementary Table 2; Materials and Methods; Fig. 6d). Similar to mRNA edits in model experiments with individual APOBECs-expressing yeast and human cells (Figs. 3d and 4a), the highest enrichment was observed with uCg motif. There was also statistically significant enrichment with another motif, uCu, belonging to the APOBEC3A-preferred uCn group, as well as with aCa motif, which falls outside this group. Although uCu and aCa enrichments were lower than uCg, there were no statistically significant differences between these 2 enrichments. Altogether, analysis of SARS-CoV-2 edits is consistent with a prominent role of APOBEC3A in editing (+)ssRNA genomes; however, other editing factors may contribute to aCa motif editing.

In summary, in all biological contexts where C to U RNA editing was identified, including mRNA editing in human cancer, hypoxia-treated monocytes, monocyte-derived macrophages mRNA, (+)ssRNA genomes of vaccine-derived viruses, and

pandemic population of SARS-CoV-2, we found the best match to the APOBEC3A-like RNA editing motifs previously identified in yeast and human cells that overexpress APOBEC3A. We therefore conclude that, in addition to being one of the most prolific DNA mutators in cancer, human APOBEC3A is also the predominant C to U global RNA editor among multiple APOBEC family enzymes.

Discussion

In this study, we used a parallel whole-genome DNA and mRNA sequencing strategy coupled with a motif-centered statistical analyses to demonstrate that full-size ORFs of APOBEC1, APOBEC3C, APOBEC3G, APOBEC3B, and APOBEC3A expressed in yeast can edit cytosines in mRNA, though to different extents. mRNA editing was evaluated in parallel with DNA mutagenesis. The latter served as an internal control of cytidine deaminase activity present in yeast. In order to make DNA mutagenesis detectable even with weaker APOBECs (APOBEC1, APOBEC3C, and dd APOBEC3G), we used yeast with a hypomorph mutation in the large subunit of RPA (*rfa1-t33*) sensitized to APOBEC mutagenesis (Dennen et al. 2024). The strongest of the editors tested in these defined experiments was APOBEC3A. We also attributed a distinct editing motif of APOBEC3A—uCg→uUg as the most enriched of the trinucleotide motifs centered around C→U edits. The diagnostic value of uCg RNA editing motif was confirmed in human cells expressing APOBEC3A. The motif preference determined via direct experimental testing and stringent statistical evaluation allowed for the development of a testable hypothesis that APOBEC3A is the primary RNA editor in vivo. To test the hypothesis, we applied our knowledge-based statistical analyses to several datasets of RNA editing human mRNAs and in (+)ssRNA viruses and found that the APOBEC3A-like editing motif was prevailing in all the biological datasets that we analyzed. Together, this allowed us to propose that though multiple APOBECs can edit RNA, APOBEC3A is the primary global RNA editor in humans.

Diagnostic features of APOBEC3A prevalence in RNA editing

Our yeast study revealed several features of C→U RNA editing that can be attributed to APOBEC3A. (i) The highest total enrichment in the uCn group of 4 trinucleotides; (ii) lack of enrichment of the remaining 12 trinucleotide motifs centered around C; and (iii) unique preference to uCg→uUg editing motif over uCa→uUa (Fig. 3). The uCg over uCa preference is unique to RNA editing and was not observed in APOBEC3A mutagenesis in the analogous ssDNA motifs tCg→tTg and tCa→tTa in yeast (Figs. 1 and 2). It remains to establish mechanisms underlying motif preferences to APOBEC3A mRNA editing; however, the combination of the 3 diagnostic features held in human cells overexpressing APOBEC3A was not observed in cells overexpressing APOBEC3B (Fig. 4a). In agreement with prior biochemical studies of APOBEC3A and APOBEC3B (Sharma and Baysal 2017; Cortez et al. 2019; Jalili et al. 2020), the APOBEC3A-expressing human cells demonstrated the strongest preference to the loop of stem-loop structures (Fig. 4b) and furthermore showed clear preference for the 3′ base of a hairpin's loop in APOBEC3A-expressing cells. Moreover, the preference of uCg motif to 3′C in the loop was the highest among 4 motifs of the uCn group (Fig. 4c).

Prevailing role of APOBEC3A in global C to U editing of transcriptome

The APOBEC3A-like editing features prevailed in mRNA editing calls from TCGA cancers and human blood cells, where the agent of C→U global RNA editing was not defined (Fig. 5). General preference to uCn motif editing in yeast was observed for APOBEC3B as well as for APOBEC3A; however, APOBEC3A-like uCg→uUg motif prevailed in cancers and in blood cells. It is surprising that APOBEC3A-like editing appears to prevail over APOBEC3B, because the latter exhibits stronger expression across human cells (Burns, Lackey, et al. 2013; Burns, Temiz, et al. 2013). It is worth noting, however, that biochemical studies (Xiao et al. 2017; Cortez et al. 2019) have shown that DNA editing by APOBEC3B was inhibited by RNA, whereas APOBEC3A-mediated DNA editing was not affected. Importantly, uCg→uUg was the most prominent peak in an extracted SBS2-RNA editing multimotif signature (Martinez-Ruiz et al. 2023; Fixman et al. 2024), which is in agreement with our individual motif-centered *P*-value-generating enrichment statistics.

Multiple variants of mRNA transcripts derived from a single gene can be generated by end processing, noncoding base modifications as well as by alternative splicing and editing events, which result in downstream modifications of protein sequence (Sharma et al. 2025). While site-specific RNA editing and alternative splicing provide additional opportunities for programmed cell differentiation, the dispersed global editing events involving mRNAs and noncoding RNAs may also carry physiological functions (Lerner et al. 2018; Christofi and Zaravinos 2019; Gassner et al. 2020; Alqassim et al. 2021; Donato et al. 2022; Pecori et al. 2022). Background of global, dispersed across multiple sites, mRNA editing would also generate a background of disordered variants of proteins with yet unknown physiological impact at the cell and organism levels (Uversky 2002; Holehouse and Kragelund 2024; Uversky 2025). Biological function assigned to global DNA and RNA editing by APOBEC enzymes is the antiviral innate immunity regulated within the large network of innate immunity interferon-stimulated genes (Harris and Dudley 2015; Babadei et al. 2024), which respond to a variety of environmental and endogenous stimuli. Thus, global mRNA editing may be an inevitable consequence of the existence of complex innate immunity system. Knowledge about relative contributions of individual APOBECs should be included in the design of future research on mechanisms generating epitranscriptome. The high input into mRNA editing capability of APOBEC3A revealed by our study suggests a testable hypothesis that can be addressed in human population. It was established that APOBEC3A–APOBEC3B deletion–fusion variant frequent in human population increases frequency of APOBEC3A-like mutagenesis (Nik-Zainal et al. 2014; Chan et al. 2015, 2019). The facilitated mutagenesis was explained by increased stability of a fusion transcript carrying complete APOBEC3A ORF with 3′ UTR of APOBEC3B mRNA (Caval et al. 2014). Based on our current study, we anticipate that the overall level of global mRNA editing as well as enrichment with uCg mRNA editing motif in individuals carrying APOBEC3A–APOBEC3B fusion variants would be greater when compared with wild-type carriers.

APOBEC3A editing can cause hypermutation in ssRNA viral genomes

Accurate detection of APOBEC-catalyzed mRNA editing by whole-transcriptome sequencing is challenging when compared with APOBEC-induced mutations in genomes of RNA viruses, because each mRNA editing event should be detected in the same molecule in which the C→U change occurred, while each mutation in viral RNA genomes can be reliably identified after propagation of mutants. Since structural constraints define exquisite specificity of APOBECs to single-strand polynucleotides, mutagenesis frequency and strand-specificity of APOBEC mutagenesis in RNA

viruses would be defined by their replication cycle. The highest frequency of mutant genomes would be expected in (+)ssRNA viruses with C→U generated in genomic (+) strand (see Figure 1 in Kockler and Gordenin 2021). Viruses with positive ssRNA genomes comprise large fraction of RNA viruses, which are distributed across multiple host species (Ward 1993) including humans carrying single or multiple APOBEC genes (Conticello et al. 2005). Indeed, the striking level of APOBEC-induced hypermutation was detected in rubella virus recovered from granulomas of patients with PID (Perelygina et al. 2019) and from immunocompetent patients (Wanat et al. 2022). Sequence comparison clearly pointed to the attenuated rubella vaccine virus being a target of hypermutation. This allowed to identify position and surrounding nucleotide context of each individual C→U mutation event and calculate enrichment with each mutation motif supported by rigorous statistical evaluation (Fig. 6a and b). Two other real-life events in which complete sequence of originating genome was known, vaccine-derived polio virus (Liu et al. 2015; Devaux et al. 2023) and pandemic isolates of SARS-CoV-2 (Klimczak et al. 2020; Ratcliff and Simmonds 2021), enabled rigorous motif enrichment calculation (Fig. 6c and d). In all cases, enrichment with APOBEC3A-like uCg motif was the strongest. This result points to APOBEC3A as the strongest candidate for the global RNA editing not only in mRNAs, but also in viruses with (+)ssRNA genomes propagating in humans. Since APOBEC3A as well as other APOBECs are under innate immunity and inflammation control, it would be important to explore the effect of innate immunity and inflammation triggers on sequence variation of (+)ssRNA virus(es) during virus persistence or propagation in humans. In support of this suggestion, C→U editing of SARS-CoV-2 RNA genomes in human cultured cells was indeed facilitated by interferons and by inflammation-related interferon INF-alfa (Nakata et al. 2023).

While our study presented evidence of prevailing role of APOBEC3A in global editing of cytosines mRNAs and in RNA viruses, this statement is based on a limited amount of data analyses. More studies are required to evaluate the level of generality of such a prevailing role. However even at the current stage, the RNA editing APOBEC3A-like uCg→uUg motif identified in our work suggests testable hypotheses about impacts of individual genotype and environmental factors on generation of RNA sequence variants in cell models and even in mRNAs of cells from individuals or in RNA viruses in humans.

Data availability

Motif-centered analyses of MAF files with mutation or edit calls were performed using the P-MACD pipeline (https://github.com/NIEHS/P-MACD) All MAF files fed into P-MACD and their derivative versions *.anz4, containing +/−20 nucleotide flanks around each base substitution are listed in Supplementary Table 2 and can be found at https://doi.org/10.5281/zenodo.18079216. Reads from paired DNA- and RNA-seq samples were submitted to SRA: Bio-project PRJNA1398811 for human samples and Bio-project PRJNA1398896 for yeast samples.

Supplemental material available at G3 online.

Acknowledgments

The contributions of the NIH authors are considered Works of the US Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the United States Centers for Disease Control and Prevention. We are grateful to Drs Paul W. Doetsch, Natalya P. Degtyareva, and Rajula Alleva Elango for advice on the manuscript and for Mr Adam B. Burkholder for help in data management.

Funding

This research was supported in part by the NIH, National Institute of Environmental Health Sciences. D.A.G. is supported by the US National Institutes of Health Intramural Research Program Project Z1AES103266; S.A.R. is supported by R01CA269784 from NCI; M.D.K. is supported by Scientific and Technological Research Council of Turkey (TUBITAK) under the grant number 123E476.

Author contributions

Conceptualization: D.A.G., Z.W.K., and S.A.R.; data curation: H.B., Z.W.K., L.P., and L.J.K.; formal analysis: H.B., L.J.K., Y.-C.H., M.D.K., and J.-L.L.; investigation: Z.W.K., M.S.D., M.E.C., and T.M.M.; methodology: D.A.G., Z.W.K., and S.A.R.; resources: L.P. and S.A.R.; software: H.B, L.J.K., and M.D.K.; supervision: D.A.G., S.A.R., and J.-L.L.; visualization: Z.W.K.; writing—original draft: Z.W.K. and D.A.G.; and writing—review and editing: Z.W.K., D.A.G., S.A.R., and H.B.

Conflicts of interest

None declared.

Literature cited

Alexandrov LB *et al.* 2013. Signatures of mutational processes in human cancer. Nature. 500:415–421. https://doi.org/10.1038/nature12477.

Alexandrov LB *et al.* 2020. The repertoire of mutational signatures in human cancer. Nature. 578:94–101. https://doi.org/10.1038/s41586-020-1943-3.

Alonso de la Vega A *et al.* 2023. Acute expression of human APOBEC3B in mice results in RNA editing and lethality. Genome Biol. 24:267. https://doi.org/10.1186/s13059-023-03115-4.

Alqassim EY *et al.* 2021. RNA editing enzyme APOBEC3A promotes pro-inflammatory M1 macrophage polarization. Commun Biol. 4:102. https://doi.org/10.1038/s42003-020-01620-x.

Azgari C, Kilinc Z, Turhan B, Circi D, Adebali O. 2021. The mutation profile of SARS-CoV-2 is primarily shaped by the host antiviral defense. Viruses. 13:394. https://doi.org/10.3390/v13030394.

Babadei O, Strobl B, Muller M, Decker T. 2024. Transcriptional control of interferon-stimulated genes. J Biol Chem. 300:. https://doi.org/10.1016/j.jbc.2024.107771.

Buisson R *et al.* 2019. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science. 364:. https://doi.org/10.1126/science.aaw2872.

Burns MB *et al.* 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 494:366–370. https://doi.org/10.1038/nature11881.

Burns MB, Temiz NA, Harris RS. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 45:977–983. https://doi.org/10.1038/ng.2701.

Caval V, Suspene R, Shapira M, Vartanian J-P, Wain-Hobson S. 2014. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3′UTR enhances chromosomal DNA damage. Nat Commun. 5:5129. https://doi.org/10.1038/ncomms6129.

Cer RZ et al. 2013. Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. Nucleic Acids Res. 41:D94–D100. https://doi.org/10.1093/nar/gks955.

Chan K et al. 2012. Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet. 8:. https://doi.org/10.1371/journal.pgen.1003149.

Chan K et al. 2015. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 47:1067–1072. https://doi.org/10.1038/ng.3378.

Chan K, Gordenin DA. 2015. Clusters of multiple mutations: incidence and molecular mechanisms. Annu Rev Genet. 49: 243–267. https://doi.org/10.1146/annurev-genet-112414-054714.

Chan K, Resnick MA, Gordenin DA. 2013. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. DNA Repair (Amst). 12:878–889. https://doi.org/10.1016/j.dnarep.2013.07.008.

Chen Z et al. 2019. Integrative genomic analyses of APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes, and immunogenicity in multiple cancer types. BMC Med Genomics. 12: 131. https://doi.org/10.1186/s12920-019-0579-3.

Cherkasova EA et al. 2002. Long-term circulation of vaccine-derived poliovirus that causes paralytic disease. J Virol. 76:6791–6799. https://doi.org/10.1128/JVI.76.13.6791-6799.2002.

Christofi T, Zaravinos A. 2019. RNA editing in the forefront of epitranscriptomics and human health. J Transl Med. 17:319. https://doi.org/10.1186/s12967-019-2071-4.

Conticello SG. 2012. Creative deaminases, self-inflicted damage, and genome evolution. Ann N Y Acad Sci. 1267:79–85. https://doi.org/10.1111/j.1749-6632.2012.06614.x.

Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol. 22:367–377. https://doi.org/10.1093/molbev/msi026.

Cortez LM et al. 2019. APOBEC3A is a prominent cytidine deaminase in breast cancer. PLoS Genet. 15:e1008545. https://doi.org/10.1371/journal.pgen.1008545.

Dananberg A, Striepen J, Rozowsky JS, Petljak M. 2024. APOBEC mutagenesis in cancer development and susceptibility. Cancers (Basel). 16:374. https://doi.org/10.3390/cancers16020374.

Dennen MS et al. 2024. Hypomorphic mutation in the large subunit of replication protein A affects mutagenesis by human APOBEC cytidine deaminases in yeast. G3 (Bethesda). 14:. https://doi.org/10.1093/g3journal/jkae196.

Devaux CA, Pontarotti P, Levasseur A, Colson P, Raoult D. 2023. Is it time to switch to a formulation other than the live attenuated poliovirus vaccine to prevent poliomyelitis? Front Public Health. 11:. https://doi.org/10.3389/fpubh.2023.1284337.

Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. 2020. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Sci Adv. 6:. https://doi.org/10.1126/sciadv.abb5813.

Domingo E, Perales C. 2018. Quasispecies and virus. Eur Biophys J. 47: 443–457. https://doi.org/10.1007/s00249-018-1282-6.

Domingo E, Perales C. 2019. Viral quasispecies. PLoS Genet. 15:. https://doi.org/10.1371/journal.pgen.1008271.

Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. Microbiol Mol Biol Rev. 76:159–216. https://doi.org/10.1128/MMBR.05023-11.

Donato L et al. 2022. Epitranscriptome analysis of oxidative stressed retinal epithelial cells depicted a possible RNA editing landscape of retinal degeneration. Antioxidants (Basel). 11:1967. https://doi.org/10.3390/antiox11101967.

Fixman B et al. 2024. Validation of the APOBEC3A-mediated RNA single base substitution signature and proposal of novel APOBEC1, APOBEC3B, and APOBEC3G RNA signatures. J Mol Biol. 436:. https://doi.org/10.1016/j.jmb.2024.168854.

Gassner FJ, Zaborsky N, Feldbacher D, Greil R, Geisberger R. 2020. RNA editing alters miRNA function in chronic lymphocytic leukemia. Cancers (Basel). 12:. https://doi.org/10.3390/cancers12051159.

Gerhauser C et al. 2018. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. Cancer Cell. 34:996–1011.e1018. https://doi.org/10.1016/j.ccell.2018.10.016.

Green AM, Weitzman MD. 2019. The spectrum of APOBEC3 activity: from anti-viral agents to anti-cancer opportunities. DNA Repair (Amst). 83:. https://doi.org/10.1016/j.dnarep.2019.102700.

Harris RS, Dudley JP. 2015. APOBECs and virus restriction. Virology. 479–480:131–145. https://doi.org/10.1016/j.virol.2015.03.012.

Holehouse AS, Kragelund BB. 2024. The molecular basis for cellular function of intrinsically disordered protein regions. Nat Rev Mol Cell Biol. 25:187–211. https://doi.org/10.1038/s41580-023-00673-0.

Hudson KM et al. 2023. Glycidamide-induced hypermutation in yeast single-stranded DNA reveals a ubiquitous clock-like mutational motif in humans. Nucleic Acids Res. 51:9075–9100. https://doi.org/10.1093/nar/gkad611.

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. 2020. Pan-cancer analysis of whole genomes. Nature. 578:82–93. https://doi.org/10.1038/s41586-020-1969-6.

Jalili P et al. 2020. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. Nat Commun. 11:2971. https://doi.org/10.1038/s41467-020-16802-8.

Kazanov MD et al. 2015. APOBEC-induced cancer mutations are uniquely enriched in early-replicating, gene-dense, and active chromatin regions. Cell Rep. 13:1103–1109. https://doi.org/10.1016/j.celrep.2015.09.077.

Kew O et al. 2002. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. Science. 296: 356–359. https://doi.org/10.1126/science.1068284.

Kew OM et al. 2004. Circulating vaccine-derived polioviruses: current state of knowledge. Bull World Health Organ. 82:16–23. https://pmc.ncbi.nlm.nih.gov/articles/PMC2585883/.

Kim K et al. 2022. The roles of APOBEC-mediated RNA editing in SARS-CoV-2 mutations, replication and fitness. Sci Rep. 12: 14972. https://doi.org/10.1038/s41598-022-19067-x.

Kim K, Shi AB, Kelley K, Chen XS. 2023. Unraveling the enzyme-substrate properties for APOBEC3A-mediated RNA editing. J Mol Biol. 435:. https://doi.org/10.1016/j.jmb.2023.168198.

Klimczak LJ, Randall TA, Saini N, Li J-L, Gordenin DA. 2020. Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. PLoS One. 15:. https://doi.org/10.1371/journal.pone.0237689.

Kockler ZW, Gordenin DA. 2021. From RNA world to SARS-CoV-2: the edited story of RNA viral evolution. Cells. 10:. https://doi.org/10.3390/cells10061557.

Krueger F. 2023. Trim Galore! https://zenodo.org/badge/latestdoi/62039322.

Lerner T et al. 2021. C-to-U RNA editing: from computational detection to experimental validation. Methods Mol Biol. 2181:51–67. https://doi.org/10.1007/978-1-0716-0787-9_4.

Lerner T, Papavasiliou FN, Pecori R. 2018. RNA editors, cofactors, and mRNA targets: an overview of the C-to-U RNA editing machinery

and its implication in human disease. Genes (Basel). 10:13. https://doi.org/10.3390/genes10010013.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Liu I-M. 2005. Breslow–Day test. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. John Wiley & Sons, Ltd. p. 560–561.
Liu Y *et al.* 2015. Circulating type 1 vaccine-derived poliovirus may evolve under the pressure of adenosine deaminases acting on RNA. J Matern Fetal Neonatal Med. 28:2096–2099. https://doi.org/10.3109/14767058.2014.979147.
Martinez-Ruiz C *et al.* 2023. Genomic-transcriptomic evolution in lung cancer and metastasis. Nature. 616:543–552. https://doi.org/10.1038/s41586-023-05706-4.
Mazina OM *et al.* 2020. Replication protein A binds RNA and promotes R-loop formation. J Biol Chem. 295:14203–14213. https://doi.org/10.1074/jbc.RA120.013812.
Mertz TM *et al.* 2025. Defining APOBEC-induced mutation signatures and modifying activities in yeast. Methods Enzymol. 713:115–161. https://doi.org/10.1016/bs.mie.2024.11.041.
Mertz TM, Collins CD, Dennis M, Coxon M, Roberts SA. 2022. APOBEC-induced mutagenesis in cancer. Annu Rev Genet. 56:229–252. https://doi.org/10.1146/annurev-genet-072920-035840.
Nakata Y *et al.* 2023. Cellular APOBEC3A deaminase drives mutations in the SARS-CoV-2 genome. Nucleic Acids Res. 51:783–795. https://doi.org/10.1093/nar/gkac1238.
Nik-Zainal S *et al.* 2014. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat Genet. 46:487–491. https://doi.org/10.1038/ng.2955.
Oh S, Buisson R. 2022. A digital PCR-based protocol to detect and quantify RNA editing events at hotspots. STAR Protoc. 3:. https://doi.org/10.1016/j.xpro.2022.101148.
Pecori R, Giorgio SD, Paulo Lorenzo J, Nina Papavasiliou F. 2022. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. Nat Rev Genet. 23:505–518. https://doi.org/10.1038/s41576-022-00459-8.
Perelygina L *et al.* 2019. Infectious vaccine-derived rubella viruses emerge, persist, and evolve in cutaneous granulomas of children with primary immunodeficiencies. PLoS Pathog. 15:. https://doi.org/10.1371/journal.ppat.1008080.
Perelygina L, Icenogle J, Sullivan KE. 2020. Rubella virus-associated chronic inflammation in primary immunodeficiency diseases. Curr Opin Allergy Clin Immunol. 20:574–581. https://doi.org/10.1097/ACI.0000000000000694.
Petljak M *et al.* 2019. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. Cell. 176:1282–1294.e1220. https://doi.org/10.1016/j.cell.2019.02.012.
Petljak M *et al.* 2022. Mechanisms of APOBEC3 mutagenesis in human cancer cells. Nature. 607:799–807. https://doi.org/10.1038/s41586-022-04972-y.
Picardi E. 2024. REDItools. https://github.com/BioinfoUNIBA/REDItools.
Ponomarev GV *et al.* 2022. APOBEC mutagenesis is low in most types of non-B DNA structures. iScience. 25:. https://doi.org/10.1016/j.isci.2022.104535.
Powell LM *et al.* 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell. 50:831–840. https://doi.org/10.1016/0092-8674(87)90510-1.
Ratcliff J, Simmonds P. 2021. Potential APOBEC-mediated RNA editing of the genomes of SARS-CoV-2 and other coronaviruses and its impact on their longer term evolution. Virology. 556:62–72. https://doi.org/10.1016/j.virol.2020.12.018.
Refsland EW, Harris RS. 2013. The APOBEC3 family of retroelement restriction factors. Curr Top Microbiol Immunol. 371:1–27. https://doi.org/10.1007/978-3-642-37765-5_1.
Roberts SA *et al.* 2012. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell. 46:424–435. https://doi.org/10.1016/j.molcel.2012.03.030.
Roberts SA *et al.* 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 45:970–976. https://doi.org/10.1038/ng.2702.
Roberts SA, Gordenin DA. 2014. eLS. Chichester: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0024941.
Saini N *et al.* 2020. Mutation signatures specific to DNA alkylating agents in yeast and cancers. Nucleic Acids Res. 48:3692–3707. https://doi.org/10.1093/nar/gkaa150.
Sakofsky CJ *et al.* 2019. Repair of multiple simultaneous double-strand breaks causes bursts of genome-wide clustered hypermutation. PLoS Biol. 17:. https://doi.org/10.1371/journal.pbio.3000464.
Sharma S *et al.* 2015. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. 6:6881. https://doi.org/10.1038/ncomms7881.
Sharma S *et al.* 2019. Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. Genome Biol. 20:37. https://doi.org/10.1186/s13059-019-1651-1.
Sharma Y *et al.* 2025. mRNA transcript variants expressed in mammalian cells. Int J Mol Sci. 26:. https://doi.org/10.3390/ijms26031052.
Sharma S, Baysal BE. 2017. Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. PeerJ. 5:. https://doi.org/10.7717/peerj.4136.
Sharma S, Patnaik SK, Kemer Z, Baysal BE. 2017. Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. RNA Biol. 14:603–610. https://doi.org/10.1080/15476286.2016.1184387.
Sharma S, Patnaik SK, Taggart RT, Baysal BE. 2016. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. Sci Rep. 6:. https://doi.org/10.1038/srep39100.
Taylor BJ *et al.* 2013. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife. 2:. https://doi.org/10.7554/eLife.00534.
Teng B, Burant CF, Davidson NO. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science. 260:1816–1819. https://doi.org/10.1126/science.8511591.
Uversky VN. 2002. Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 11:739–756. https://doi.org/10.1110/ps.4210102.
Uversky VN. 2025. Functional diversity of intrinsically disordered proteins and their structural heterogeneity: protein structure-function continuum. Prog Mol Biol Transl Sci. 211:1–15. https://doi.org/10.1016/bs.pmbts.2024.11.006.
Wanat KA *et al.* 2022. Association of persistent Rubella virus with idiopathic skin granulomas in clinically immunocompetent adults. JAMA Dermatol. 158:626–633. https://doi.org/10.1001/jamadermatol.2022.0828.
Ward CW. 1993. Progress towards a higher taxonomy of viruses. Res Virol. 144:419–453. https://doi.org/10.1016/S0923-2516(06)80059-2.
Way J. 2022. Picard. https://github.com/broadinstitute/picard/releases/tag/2.26.11
Wong L, Vizeacoumar FS, Vizeacoumar FJ, Chelico L. 2021. APOBEC1 cytosine deaminase activity on single-stranded DNA is suppressed by replication protein A. Nucleic Acids Res. 49:322–339. https://doi.org/10.1093/nar/gkaa1201.
Xiao X *et al.* 2017. Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. Nucleic Acids Res. 45:7494–7506. https://doi.org/10.1093/nar/gkx362.

Yan D et al. 2010. Emergence and localized circulation of a vaccine-derived poliovirus in an isolated mountain community in Guangxi, China. J Clin Microbiol. 48:3274–3280. https://doi.org/10.1128/JCM.00712-10.

Yang C-F et al. 2005. Intratypic recombination among lineages of type 1 vaccine-derived poliovirus emerging during chronic infection of an immunodeficient patient. J Virol. 79: 12623–12634. https://doi.org/10.1128/JVI.79.20.12623-12634.2005.

Yang Y et al. 2025. An expanding universe of mutational signatures and its rapid evolution in single-stranded RNA viruses. Mol Biol Evol. 42:. https://doi.org/10.1093/molbev/msaf009.

Zhang Y, Park C, Bennett C, Thornton M, Kim D. 2021. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. Genome Res. 31:1290–1295. https://doi.org/10.1101/gr.275193.120.

Editor: N. Rhind