Molecular Therapy
Nucleic Acids
Original Article

# MultiEditR: The first tool for the detection and quantification of RNA editing from Sanger sequencing demonstrates comparable fidelity to RNA-seq

Mitchell G. Kluesner,[1,2,3,4,5,6] Rafail Nikolaos Tasakis,[7,8] Taga Lerner,[7,8] Annette Arnold,[7] Sandra Wüst,[9] Marco Binder,[9] Beau R. Webber,[1,2,3,4,10] Branden S. Moriarity,[1,2,3,4,10] and Riccardo Pecori[7,10]

[1]Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455, USA; [2]Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55455, USA; [3]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA; [4]Stem Cell Institute, University of Minnesota, Minneapolis, MN 55455, USA; [5]University of Washington School of Medicine, Seattle, WA 98195, USA; [6]Medical Scientist Training Program, University of Washington, Seattle, WA 98195, USA; [7]Division of Immune Diversity, Program in Cancer Immunology, German Cancer Research Centre (DKFZ), 69120 Heidelberg, Germany; [8]Faculty of Biosciences, Heidelberg University, 69120 Heidelberg, Germany; [9]Research Group "Dynamics of Early Viral Infection and the Innate Antiviral Response," Division Virus Associated Carcinogenesis (F170), German Cancer Research Centre (DKFZ), 69120 Heidelberg, Germany

**We present MultiEditR (Multiple Edit Deconvolution by Inference of Traces in R), the first algorithm specifically designed to detect and quantify RNA editing from Sanger sequencing (z.umn.edu/multieditr). Although RNA editing is routinely evaluated by measuring the heights of peaks from Sanger sequencing traces, the accuracy and precision of this approach has yet to be evaluated against gold standard next-generation sequencing methods. Through a comprehensive comparison to RNA sequencing (RNA-seq) and amplicon-based deep sequencing, we show that MultiEditR is accurate, precise, and reliable for detecting endogenous and programmable RNA editing.**

## INTRODUCTION

RNA editing is the most abundant post-transcriptional modification in messenger RNA (mRNA),[1] with two predominant types of editing: cytidine-to-uridine editing (C-to-U) by the APOBEC family of enzymes, and adenosine-to-inosine editing (A-to-I) by the ADAR family of enzymes. RNA editing has implications in a variety of biologically processes, particularly among those involved in neural physiology, immunity, and oncogenesis.[2,3] Importantly, the recent development of programmable RNA base-editing technologies presents the possibility to correct pathogenic mutations at the RNA level, opening important therapeutic scenarios.[4] For both endogenous and programmable RNA editing, the accurate and precise detection as well as quantification of editing is essential.

Current identification and quantification of endogenous RNA editing relies on RNA-sequencing (RNA-seq) data analyzed by several different algorithmic approaches.[5,6] Although these approaches are robust, they are often complicated by genomic sequence polymorphisms, sequencing errors, and-or low coverage of certain genomic regions. This leads to the necessity of routine validation and quantification of RNA editing sites by Sanger sequencing, and by bacterial colony sequencing of subcloned polymerase chain reaction (PCR) amplicons.[1,7]

Meanwhile, identification and quantification of programmable RNA editing is mainly accomplished by Sanger sequencing.[8–11] To validate the existence of programmed RNA edits, the first step is to evaluate the editing efficiency of the targeted base (on-target editing), and the second step is to evaluate the presence of possible undesired editing sites along the same transcript (off-target editing). This part of the process, independently from the RNA base-editing method used, represents a time-consuming step, in which a quick and inexpensive evaluation of on-target and off-target editing at the transcript level is needed to define the best experimental settings to use.

Despite the widespread use of Sanger sequencing, no program exists specifically for the quantification of RNA editing, leaving some to creatively, yet not optimally, measure the height of peaks from Sanger sequencing traces using image analysis software such as Adobe Illustrator.[12] Although some tools exist for the quantification of DNA base editing from Sanger sequencing,[13] they are not able to detect and quantify multiple sites simultaneously within the same Sanger trace. This feature is essential to evaluate on-target and off-target programmable RNA editing, and it is also crucial in the context of endogenous RNA editing, where editing sites are often in a cluster, leading to hyper-edited transcript regions that can be missed by standard RNA editing analysis of RNA-seq data.[1,14] Finally, the detection and quantification capabilities of a Sanger based approach has yet to be benchmarked against

Received 8 January 2021; accepted 13 July 2021;
https://doi.org/10.1016/j.omtn.2021.07.008.

[10]These authors contributed equally

**Correspondence:** Riccardo Pecori, PhD, Division of Immune Diversity, Program in Cancer Immunology, German Cancer Research Centre (DKFZ), 69120 Heidelberg, Germany.
**E-mail:** r.pecori@dkfz.de

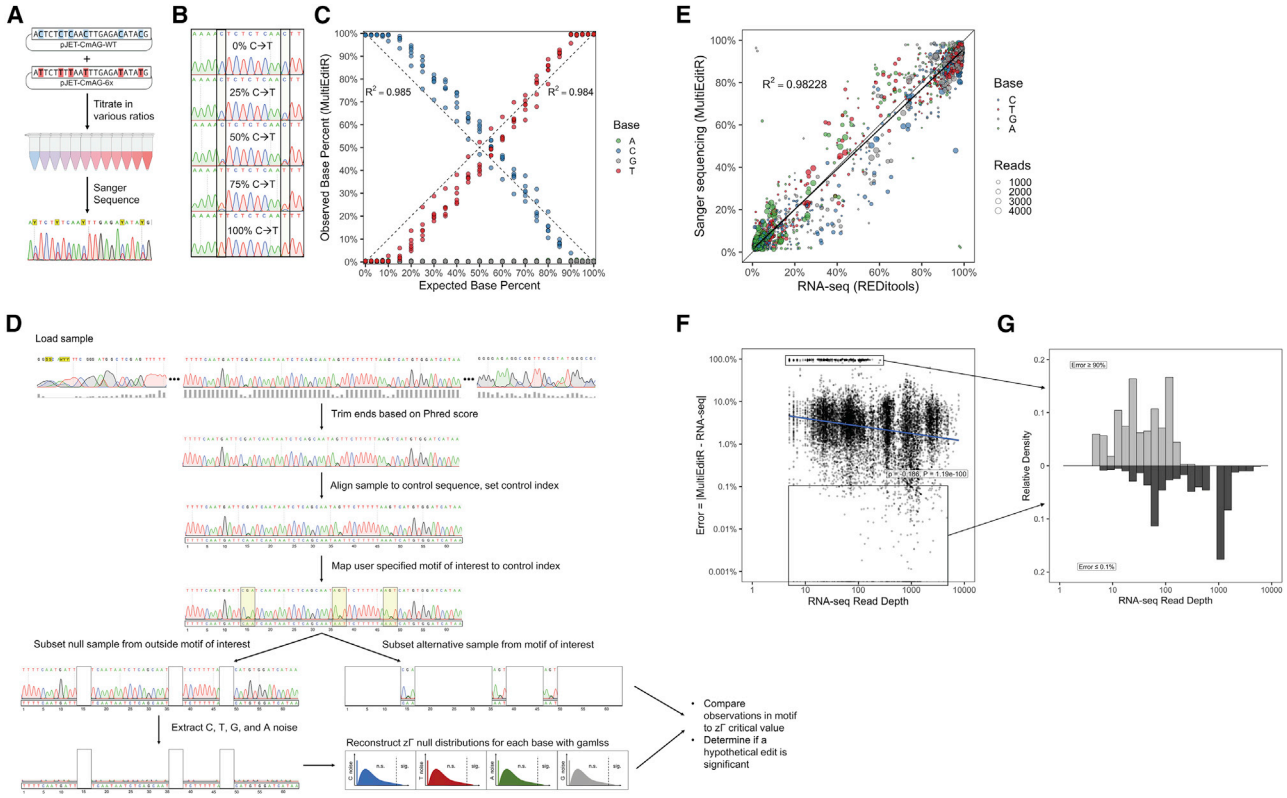



**Figure 1. Development and initial assessment of MultiEditR**

(A) Experimental scheme for the plasmid titration experiment. (B) Representative chromatograms from C-to-T titration experiment showing a change in peak height at two sites. (C) Titration of pJET-CmAG-WT with pJET-CmAG-6x sequenced with the reverse primer; the coefficient of determination ($R^2$) was calculated relative to the identity line. n = 6 editing sites per chromatogram. (D) Diagram of MultiEditR algorithm showing end trimming, sample alignment, motif mapping, zero adjusted gamma (zΓ) distribution generation, and null hypothesis significance testing. The MultiEditR algorithm can be applied to any motifs, WT base, and edited base identity. (E) Scatterplot of measurements of editing from Sanger sequencing by MultiEditR against measurements of editing from RNA-seq by REDItools. Coefficient of determination ($R^2$) represents regression to the identity line. The overlaid black line is the linear model of best fit. Dot size is proportional to RNA-seq read coverage at the base of interest. (F) Absolute value of error in MultiEditR measurements by RNA-seq read depth. Population at the top of the graph with low read depth represents bases with >90% error in measurements. Data were analyzed by Spearman's rank-order correlation test. (G) Mirrored histogram showing populations with high (≥90%) error and low (≤0.1%) error in MultiEditR measurements as a function of RNA-seq read depth.

next-generation sequencing (NGS) methods. To meet these needs we developed multiple edit deconvolution by inference of traces in R (MultiEditR) (z.umn.edu/multiedit r), a program with a web interface that provides accurate and cost-effective detection and quantification of RNA editing from Sanger sequencing that yields comparable results to NGS methods.

## RESULTS

To first determine whether successive base edits could be accurately quantified from Sanger sequencing, we titrated two plasmids that differed by six C-to-T mutations (Supplemental materials and methods; Figure S1A) and subjected the titrations to Sanger sequencing (Figure 1A). Analysis of the traces from both the forward and reverse direction showed that the percent height of the mixed peaks of interest yielded well-fit linear regressions compared to the expected titrated percent for both C-to-T and G-to-A titrations (Figures 1B and 1C; Figures S1B–S1E). Encouraged by the results we adapted the EditR algorithm, which we previously developed for analyzing CRISPR-Cas9 DNA base editing,[13] for the detection and quantification of multiple RNA edits from Sanger sequencing, which we termed MultiEditR (Figure 1D; Figure S2). To compare the performance of MultiEditR with standard methods for RNA editing detection, we first generated two knockout (KO) cell lines for ADAR1 and APOBEC1 (Figure S3). Next, we performed RNA-seq on RNA from both wild-type (WT) and KO cell lines and analyzed the data with REDItools,[15,16] a well-established tool for detection of RNA editing from RNA-seq. From the same samples several regions within different transcripts (Table S1) were PCR amplified, Sanger sequenced, and analysis of these traces was performed by MultiEditR. An initial comparison of MultiEditR to RNA-seq showed that although the central tendency of MultiEditR measurements were accurate, there was substantial error relative to the RNA-seq benchmark (Figure 1E; Figure S4). Despite the high coverage of our RNA-seq experiments (62–88 million mapped reads per sample), this error appeared to be influenced by the read depth per base of the RNA-seq dataset (Figures 1F and 1G; Figure S4), which is consistent with findings by others.[7]

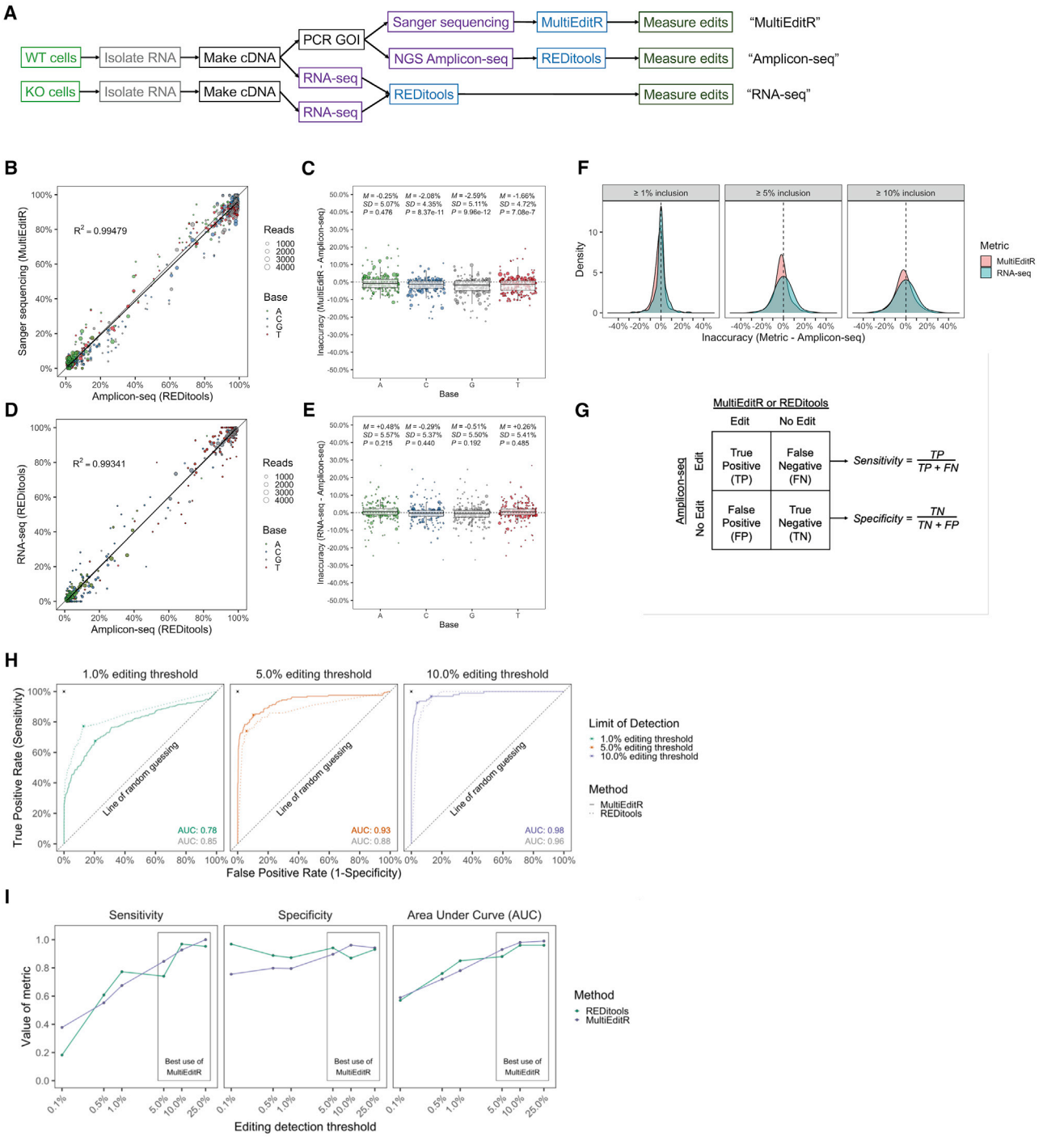



**Figure 2. Quantification and detection of endogenous RNA editing by MultiEditR**

(A) Workflow for generation of sequencing data for three-way comparison of MultiEditR, RNA-seq, and Amplicon-seq. RNA-seq from WT cell lines was compared to RNA-seq from paired APOBEC1 or ADAR1 KO cell lines to identify putative editing sites, which were PCR amplified from WT cell line cDNA. Amplicons were then subjected to Sanger sequencing and NGS deep sequencing in parallel and analyzed by MultiEditR or REDItools, respectively. (B–E) Comparison of MultiEditR and RNA-seq measurements of editing to high coverage Amplicon-seq measurements. All data are filtered on between 1% and 99% editing as measured from Amplicon-seq data. Coefficients of determination ($R^2$) represent regression to the identity line. Dot size is proportional to read coverage at the base of interest in the RNA-seq data. M = mean, SD = standard deviation. p values (P) were calculated by Student's one-sample two-tailed t test. Boxplot center lines represent the median, box limits represent the upper and lower quartiles, and

*(legend continued on next page)*

To address the potential issue that the low coverage of the RNA-seq dataset was introducing error in assessing the accuracy of MultiEditR, we performed a three-way, matched comparison of endogenous RNA editing quantification and detection within several transcripts from two cell lines (Table S2) using MultiEditR, RNA-seq, and high coverage amplicon based NGS (termed, Amplicon-seq) (Figure 2A). The direct comparison of MultiEditR to Amplicon-seq demonstrated that MultiEditR is on average accurate relative to Amplicon-seq with small significant inaccuracies among edits measured from C, G, and T bases (range of M: −1.66% to −2.59%, $p < 0.001$), and non-significantly different for edits measured from A bases (M = −0.25%, $p = 0.476$) (Figures 2B and 2C). These small inaccuracies may be attributable to peak-ratio bias, which is a well-known aspect of Sanger sequencing.[17,18] In comparison, RNA-seq benchmarked against Amplicon-seq exhibited no significant inaccuracies across all bases (Figures 2D and 2E; Figure S5). However, RNA-seq exhibited a greater standard deviation than did MultiEditR for all bases. Importantly, we found that MultiEditR was measured as more precise when benchmarked against Amplicon-seq as opposed to RNA-seq (Figures 2B and 2C; Figures S5E–S5I), confirming that the observed error in MultiEditR detection relative to the RNA-seq (Figure S4) was indeed due to low coverage in some regions of the RNA-seq dataset. Collectively, these results indicate that compared to Amplicon-seq, while the quantification of RNA editing by RNA-seq is more accurate, MultiEditR is more precise than RNA-seq, particularly when looking at editing events above 5% editing (Figure 2F).

Next, we wanted to assess how MultiEditR and REDItools analysis of RNA-seq[15,16] perform in the detection of edits (Figure 2G). Using the MultiEditR p value calculated from the zero-adjusted gamma distribution null hypothesis significance test (Figure 1D), as well as the p value from the REDItools Fisher exact test as classifier values, we performed a receiving operating characteristic (ROC) curve analysis. When including edits that were called 1% or greater by Amplicon-seq, RNA-seq performed modestly better at detecting edits than did MultiEditR in terms of sensitivity, specificity, and area under the curve (AUC), among other metrics (Figure 2H; Figure S6). However, when examining edits that were called 5% or greater by Amplicon-seq, MultiEditR performed better than RNA-seq (Figure 2I). Furthermore, running ROC curve analyses across a range of editing detection thresholds suggests that based on the measured sensitivity, specificity, and AUC, the optimal use of MultiEditR is for detecting editing events $\geq$5% (Figure 2I).

Last, we wanted to assess the utility of MultiEditR in application to biologically relevant problems. Roth et al.[19] proposed the Alu editing index (AEI, here as $EI_{NGS}$) as an index for the quantification of global RNA editing. Here, we apply a similar approach to develop the MultiEditR editing index (MEI) as a local editing index across the Sanger trace (Figure 3A). Using our three-way, matched dataset we found that the MEI is moderately correlated with the lower read depth per base RNA-seq $EI_{NGS}$ ($r = 0.558$, $p = 3.48e{-7}$), and it is well correlated with the higher read depth per base Amplicon-seq $EI_{NGS}$ ($r = 0.812$, $p = 1.39e{-5}$) (Figures 3B and 3C), further showing an effect of read depth in RNA editing detection and quantification from RNA-seq data. Using the MEI we wanted to investigate the effect of adding a nuclear localization signal (NLS) on the specificity of the programmable 4λN-ADAR2$_{DD}$ A-to-I editing system as previously published[20] (Figure 3D). Using a fluorescent reporter, we were able to directly compare editing rates to a functional readout via flow cytometry, as well as measuring editing across the transcript with the MEI (Figure 3E). We found that MultiEditR measurements of editing agreed well with flow cytometry values (Figure 3F). Additionally, using a normalized metric of percent editing of the target base (on-target editing) divided by the MEI (off-target editing), we recapitulated results that addition of an NLS to the 4λN-ADAR2$_{DD}$ system improves editing specificity[20] (Figure 3G; Figure S7). Last, we wanted to determine whether MultiEditR could be used to similarly quantify CRISPR-Cas9 DNA base editing using data previously published by our group from work using base editing to disrupt genes via splice-site targeting[21] (Figure 3H). Base editing efficiency measured by MultiEditR compared to CRISPR-DAV analysis of Amplicon-seq,[22] as well as flow cytometry, yielded strong coefficients of determination at both the DNA ($R^2 = 0.97$, Figure 3I) and protein level ($R^2 = 0.849$, Figure 3J).

For the best use of MultiEditR, we recommend designing primers to amplify a 350- to 700-bp amplicon to allow for a long enough sequence to construct null distributions for edit detection. Additionally, we recommend the use of one-step RT-PCR kits, over standard cDNA synthesis kits, to exclusively generate cDNA from the transcript of interest. Following amplification, a column-based PCR purification step is typically sufficient to ensure clean sequencing results. For detecting edits above 5%, we recommend using $p = 0.001$ for applications where false positives are strongly disfavored, and $p = 0.01$ when an increase in sensitivity is valued over a loss in specificity (Figure S6). Finally, for the best accuracy and precision of MultiEditR we recommend measuring the edit from the T or A base (Figure 2C). Last, due to the sensitivity of MultiEditR, we do not recommend using it to detect or measure edits that are below 5% due to the baseline noise in Sanger traces. For applications where high accuracy and a low limit of edit detection is paramount, we recommend using Amplicon-seq.

## DISCUSSION

Collectively, we developed MultiEditR, the first algorithm specifically designed to detect and quantify multiple RNA editing sites in a single

whiskers define the 1.5× interquartile range. (F) Kernel plots of the inaccuracy of MultiEditR and RNA-seq relative to Amplicon-seq. Plots are faceted by 1%, 5%, or 10% editing inclusion cutoffs based on the editing measured in Amplicon-seq. (G) Confusion matrix that defines the sensitivity and specificity of the edit detected by MultiEditR or REDItools, against the true state of editing, defined by Amplicon-seq. (H) ROC curves of MultiEditR and REDItools at 1%, 5%, and 10% editing cutoffs, where an Amplicon-seq event is a true edit. MultiEditR is shown as a solid curve, while REDItools is shown as a dotted curve. Crossed points along the curves represent the closest geometric distance of the curve to the upper-left error-free point at 0% and 100%. (I) ROC analysis metrics for MultiEditR and REDItools at editing thresholds from 0.1% to 25% editing inclusion cutoff. Sensitivity and specificity are defined in (G); area under the curve (AUC) represents the integral of the ROC curve from 0% to 100%.

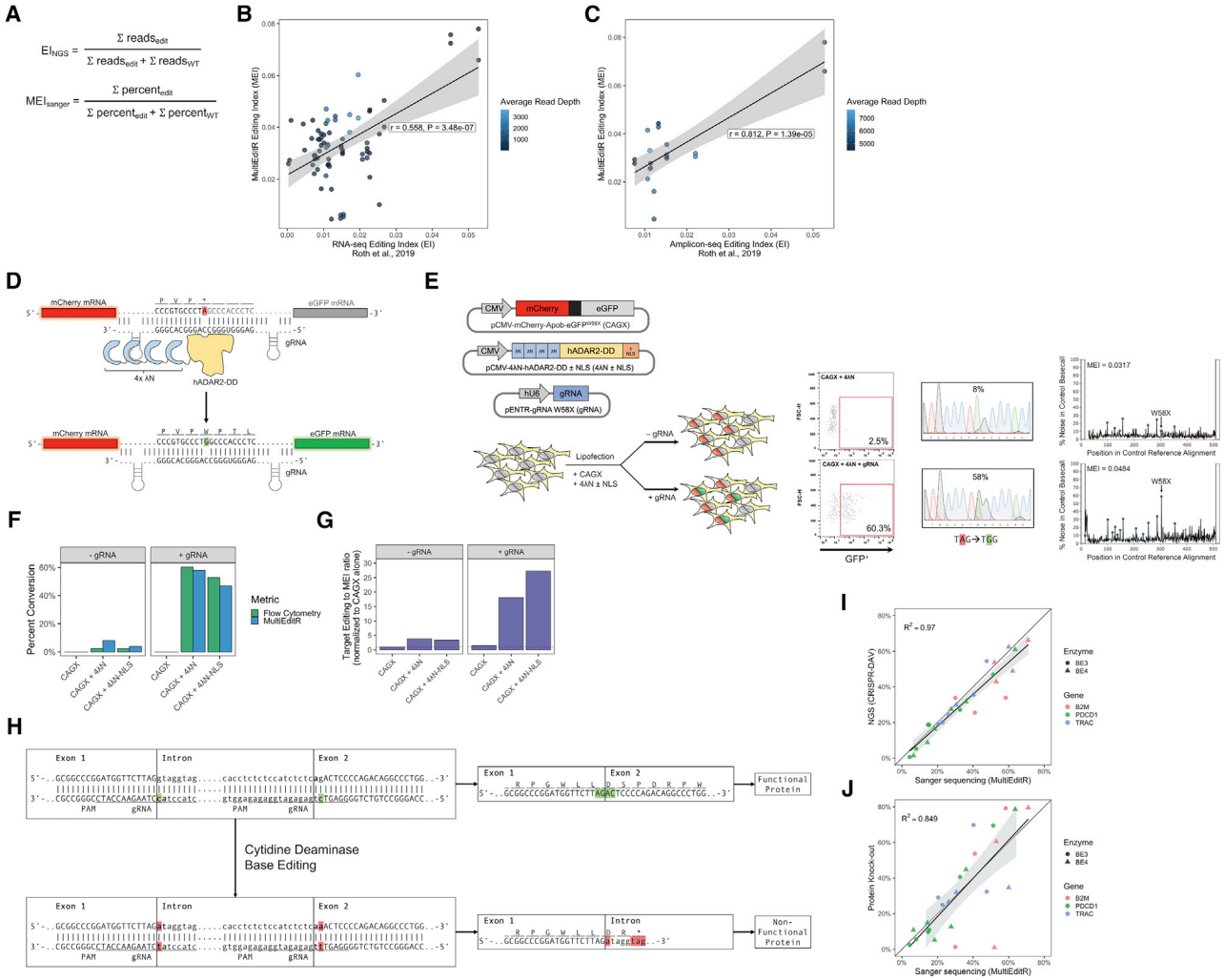


**Figure 3. Applications of MultiEditR to RNA and DNA base editing**

(A) Formula for MultiEditR editing index (MEISanger) and Alu editing index from Roth et al.[19] calculated from NGS data (EINGS). (B) Correlation between MEI and RNA-seq EI shaded by average read depth of RNA-seq. (C) Correlation between MEI and Amplicon-seq EI shaded by average read depth of Amplicon-seq. (D) Diagram of 4λN-hADAR2-DD RNA editing system. (E) EGFP reactivation experiment using the 4λN-DD system. Plasmids encoding gRNA, 4λN-hADAR2-DD, and mCherry-Apob-EGFPW58X (CAGX) were transfected into HEK293T cells. (F) Reactivation of EGFP measured via flow cytometry and MultiEditR. (G) Ratio of target X58W editing to non-desired events, measured by the ratio of target editing to the MEI. Results are normalized to the CAGX alone negative control. Flow cytometry plots, MultiEditR plots, and analysis parameters are available in Figure S7. (H and I) DNA base editing quantification of rAPOBEC1-BE3 and rAPOBEC1-BE4 in human T cells by MultiEditR compared to Amplicon-seq NGS analyzed by CRISPR-DAV. (H and J) Comparison of MultiEditR to protein KO efficiency as measured by FACS in human T cells; data were analyzed from Webber et al.[21]

trace of Sanger sequencing and we performed a comprehensive comparison with NGS methods to evaluate the performance of the tool. MultiEditR showed higher precision in RNA editing detection than did RNA-seq, particularly when looking at editing events above 5% editing, but with the cost of lower accuracy. Furthermore, in the context of RNA programmable editing, the capability of MultiEditR to detect multiple edits simultaneously and the MEI allow for a quick and inexpensive evaluation of on-target and off-target editing at the transcript level (Figure 3), a crucial aspect to define the best experimental conditions for mutation correction at the RNA level (e.g., choosing an optimal guide RNA).

Finally, we showed that MultiEditR can be employed for a variety of nucleic acid editing applications, including endogenous RNA editing, targeted programmable RNA editing, off-target RNA editing, and DNA base editing. The flexibility of the MultiEditR algorithm allows our approach to be readily applied to other applications that involve the change of one base species to another, such as that involved in bisulfite sequencing for identifying methylation, or more recently RNA polymerases and reverse transcriptases recoding various RNA modifications with distinct fidelity.[23] Overall, we predict that MultiEditR and the comparisons detailed in this study will have immediate use to the RNA editing community,

but also more broadly to the many burgeoning fields studying nucleic acid modifications.

## MATERIALS AND METHODS

### Plasmids

The mCherry-mApob-EGFP plasmid (CmAG) was obtained by substituting the human *APOB* with mouse *Apob (*mApob) in the original plasmid mCherry-APOB-EGFP[24] (kind gifts from Dr. Silvestro Conticello, Florence, Italy). mCherry-APOB-EGFP was digested with HindIII-SmaI and a PCR fragment of mouse *Apob* (467 bp from RNA of jejunal epithelial cells from the small intestines of C57BL/6 mice,[25] oligonucleotides [oligos] #1–2) was inserted into the plasmid using NEBuilder HiFi DNA assembly master mix (NEB). The mouse APOBEC1 expression vector (pCMV APOBEC1) was a kind gift from Dr. Dewi Harjanto (Laboratory of Lymphocyte Biology, The Rockefeller University). The mouse RBM47 expression vector (pCMV RBM47) was obtained by inserting a PCR fragment containing the coding sequence of mouse RBM47 (transcript variant 4, mRNA sequence ID: GenBank: NM_001291226.1) into the mCherry-Apob-EGFP cut with NheI-BsrGI. The amplification was done using oligos #3–4 on RNA of jejunal epithelial cells from the small intestines of C57BL/6 mice[25] and the cloning with NEBuilder HiFi DNA assembly master mix (NEB).

LentiCRISPRv2 was a gift from Dr. Feng Zhang (Addgene, plasmid #52961; http://addgene.org/52961; RRID:Addgene_52961).[26] DNA oligos #5–6 were cloned into this plasmid following the "lentiCRISPRv2 and lentiGuide oligo cloning protocol" (Addgene plasmid #52961) to generate lenti-CRISPR-ADAR1 exon 4 (from Pestal et al.[27]). As a non-editing transduction control, lenti-CRISPR-NT (Lenti-NT) was cloned accordingly using oligos #7–8. pCMV-DR8.91 (coding for HIV gag-pol) and pMD2.G (encoding the VSV-G glycoprotein) were kind gifts from Prof. Didier Trono (Lausanne, Switzerland). pSpCas9(BB)-2A-GFP (PX458) was a gift from Feng Zhang (Addgene plasmid #48138; http://addgene.org/48138; RRID:Addgene_48138).[28] The plasmid was digested with BsbI (NEB) and dephosphorylated with a RAPID DNA Dephos and Ligation kit (Roche). Oligos #9–12 are all 5′ phosphorylated. The oligo pairs #9–10 and #11–12 containing complementary sequences were annealed to each other and then ligated to the dephosphorylated PX458 to generate plasmids PX458-iv-single guide RNA (sgRNA)-A1_11 (cutting in exon 4) and PX458-iv-sgRNA-A1_39 (cutting in exon 5), respectively.

The mCherry-APOB-EGFP W58X plasmid (CAGX) was obtained by site-directed mutagenesis using oligos #13–14 and QuikChange Lightning site-directed mutagenesis kit (Agilent, #210518) on the original plasmid mCherry-APOB-EGFP.[24] 4λN-DD E488Q ADAR2 (4λN) and U6 pENTR gRNA vectors were a kind gift of Dr. Joshua Rosenthal (University of Chicago).[8] The NLS version of 4λN plasmid (4λN-NLS) was created by adding the c-*myc* NLS to the C-terminus of 4λN. The U6 pENTR gRNA vector was linearized by PCR using oligos #15–16 and Q5 high-fidelity DNA polymerase (NEB). The sequence containing the gRNA to induce specific A-to-G editing on the W58X of CAGX (Rosenthal fashion[8]) was inserted into the linearized pENTR using oligo #17 and NEBuilder HiFi DNA assembly master mix (NEB).

### Cell lines

A549 cells (A-549, RRID:CVCL_0023, DKFZ Germany) were cultured at 37°C, 5% $CO_2$ in high-glucose DMEM (Sigma) supplemented with 10% fetal bovine serum (FBS, PAN Biotech) and penicillin/streptomycin (Sigma). HEK293T cells (ATCC-CRL-3216) were cultured at 37°C, 5% $CO_2$ in high-glucose DMEM (Sigma) supplemented with 5% FBS (PAN Biotech) and penicillin/streptomycin (Sigma). RAW 264.7 cells (ATCC TIB-71) were cultured at 37°C, 5% $CO_2$ in high-glucose DMEM (Sigma-Aldrich) supplemented with 5% endotoxin low FBS (Sera Pro FBS, PAN Biotech), 1% glutamine, and 1% penicillin/streptomycin (Sigma). The cell lines were regularly tested for mycoplasma contamination in our facility Multiplexion (F020, DKFZ) (https://www.multiplexion.de)

### Generation of A549 ADAR1 KO cell line

Lenti-CRISPR-ADAR1 exon 4 or NT in combination with pCMV-DR8.91 and pMD2.G were calcium phosphate transfected in HEK293T cells for lentiviral particle production (ratio 3:1:3). 48–72 h after transfection, cell-free supernatant was harvested and used for transduction of A549 cells. The transduced cells were selected with puromycin (1 μg/mL). Immediately after the selection control (non-transduced A549) died, limiting dilution in 96-well plates was performed for ADAR1 KOs (0.5 cell/well) and clonality was validated by visual inspection with a microscope; the Lenti-NT control was kept polyclonal. KO of ADAR1 was validated by western blot (anti-human ADAR1 [D7E2M] rabbit monoclonal antibody [mAb], Cell Signaling Technology, cat. #14175). Two clones, numbers 5 and 7, resulted in a completely abolished ADAR1 (p110 and p150) expression (Figure S3A). For further experiments we used only clone #5.

### Generation of the RAW 264.7 APOBEC1 KO cell line

PX458-iv-sgRNA-A1_11 and PX458-iv-sgRNA-A1_39 plasmids were co-transfected using the Amaxa cell line Nucleofector kit V (Lonza) into RAW 264.7 cells following the manufacturer's protocol for RAW 264.7 cells and a Nucleofector 2b device (Lonza). 48 h post-transfection GFP-positive cells were single cell sorted into 96-well plates and clonality was validated by visual inspection with a microscope. Clones were screened by amplifying targeted regions from genomic DNA (produced by a High Pure PCR template preparation kit [Roche]) using oligos #20–21 and #22–23 and then Sanger sequencing. This was followed by additional cloning of amplified regions using a CloneJET PCR cloning kit according to the manufacturer's instructions and transforming DH5α bacteria with ligated product. Ten resultant bacteria colonies were sent for sequencing to determine genetic changes to the targeted region. One clone that was subsequently used contained in the region targeted by PX458-iv-sgRNA-A1_39 either a 1-bp deletion or a 2-bp deletion. KO was further confirmed by RT-PCR (using a One-step RT-PCR kit [QIAGEN]) amplification of B2m the 3′ UTR region from extracted RNA defined by oligos #24–25 known to be edited and determining

absence of editing compared to the amplified region from the parental cells (Figure S3A).

### RNA extraction, DNase treatment, and RT-PCR

RNA was extracted using an RNeasy mini kit (QIAGEN) and treated with DNase (Turbo DNA-free kit, Invitrogen). All of the PCRs on RNA were performed with gene-specific primers (Table S3) and a One-step RT-PCR kit (QIAGEN). Primers were designed using Primer-BLAST[29] or AmplifX 2.0.7 (https://inp.univ-amu.fr/en/amplifx-manage-test-and-design-your-primers-for-pcr) to obtain 350- to 700-bp PCR amplicons. At this stage, PCR clean up usually is sufficient, however, gel extraction is required when amplification results in multiple bands (Macherey-Nagel NucleoSpin gel and a PCR clean-up kit was used). The fragments were then subjected to Sanger sequencing (Eurofins Genomics, GATC services, Germany, or Microsynth, Switzerland) and the resulting .ab1 files were analyzed by MultiEditR.

### Titration experiments

For the C-to-U editing HEK293T cells were transfected with CmAG (50 ng), APOBEC1 (200 ng), and RBM47 (200 ng) expression vectors or CmAG (50 ng) alone. For transfection we used a mix of plasmid DNA and polyethylenimine (PEI) in an approximately 1:4 ratio (450 ng of DNA/2 μg of PEI). 72 h after transfection RNA was extracted and cDNA was amplified from *Apob* (using oligos #26 and #2). This allowed us to obtain *Apob* fragments heavily edited or not edited, respectively. These two fragments were cloned into a CloneJET PCR cloning kit (Thermo Scientific), and several colonies were screened by sequencing. From this screening we obtained two pJET vectors containing *Apob* with no editing (pJET-CmAG-WT) and six edited sites (pJET-CmAG-6x). These two vectors were then mixed together in titrated amounts from 0% to 100% and subjected to capillary Sanger sequencing with universal primers pJET1.2 forward and reverse.

### RNA-seq

RNA-seq libraries were prepared in duplicate from A549 WT and ADAR1 KO clone 5 (Figure S3) and in triplicates from RAW 264.7 WT and RAW 264.7 APOBEC1 KO. Total RNA was extracted from 10,000,000 cells in duplicate (A549 WT and ADAR1 KO) or triplicate (RAW 264.7 and RAW 264.7 APOBEC1 KO each from separate plates). RNA was extracted using an RNeasy mini kit (QIAGEN) and then treated with turbo DNase (Life Technologies). RNA concentration and integrity were determined by a Qubit 4 (Thermo Fisher Scientific) using the Qubit RNA BR assay kit or Qubit XR assay kit and the Qubit RNA IQ kit (Thermo Fisher Scientific). 1 μg of RNA was processed with a Kapa mRNA HyperPrep kit for Illumina platforms (Kapa Biosystems, Roche) and KAPA single-indexed adaptor kit for Illumina platforms (Kapa Biosystems, Roche).

Libraries were sequenced with the Illumina HiSeq 2000 v4 technology, generating 125-nt paired-end reads. Adapters were trimmed using the Trim Galore software (https://github.com/FelixKrueger/TrimGalore). Before and after trimming we evaluated the RNA-seq quality with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality control, including per-base quality, duplication levels, and over-representative sequences, passed all of the checkpoints. The RNA-seq reads (~90 million raw reads per sample) were then aligned to hg19 (A549 data) or mm10 (RAW 264.7 data) reference genomes (publicly available by the UCSC genome browser) using the STAR aligner (v2.6.0a, https://github.com/alexdobin/STAR,[30]) with default settings, resulting in 85% uniquely mapped reads (~75 million mapped reads per sample). Potential PCR duplicates (<0.25%) were removed from the aligned reads using the MarkDuplicates function from Picard tools (http://broadinstitute.github.io/picard).

The aligned RNA-seq files (.bam files) were sorted and indexed with SAMtools (https://github.com/samtools/samtools). The sorted .bam files were used as an input for the REDItoolsDnaRna.py script, part of the REDItools suite (v1.0.4[15,16]). REDItoolsDnaRna.py performs a comparative position-per-position analysis in parallel between an RNA and a DNA .bam file, so as to eliminate variants on the RNA, the signal of which derives from the genomic DNA. For this analysis, however, we used RNA-seq from A549 ADAR1 WT and ADAR1 KO, or RAW 264.7 WT and RAW 264.7 APOBEC1 KO. Sequences from KOs were used as background in order to identify the editing events for which ADAR1 or APOBEC1 is responsible. The options we set for a genomic position to be considered for variant calling required minimum coverage of five reads, with at least three reads supporting the editing event, and a minimum FastQ offset value of 33, per base. The aforementioned settings were used to run the analysis per pair and a specific gene's coordinates with the option -Y. Genes and their coordinates are listed in Table S1.

Significance testing for detecting edits from the RNA-seq data was performed according to the REDItools algorithm.[15,16] Briefly, p values were calculated for each mapped position using a Fisher's exact test. The null column was composed of the sum of reference reads (WT, i.e., A, C, G, or T) and the sum of edited reads for each reference base (edited, i.e., A-to-G, C-to-T, G-to-A, or T-to-C) across the entire RNA-seq experiment, yielding a different null contingency for each reference base. The observed number of reference (WT) reads and alternative (edited) reads at each site of interest were then compared to the null contingencies via the Fisher's exact test, generating a p value.

### Amplicon-seq

Deep amplicon NGS-seq data were generated for a set of transcripts (Table S2), each one of which was amplified from RNA using one-step RT-PCR (QIAGEN) using primers containing an adaptor sequence needed for NGSelect Amplicons 2nd PCR service by Eurofins Genomics (all of the primers used for this experiment are listed in Table S3, #57-76). Adaptor-trimmed fastq files delivered (paired end) were merged for all of the transcripts per pair. Reads were mapped to hg19 and mm10 reference genomes (UCSC) with the STAR aligner (v2.7.3a). Aligned data (.bam files) were sorted with Samtools (v1.9), and the bamUtils (v1.0.14) clipOverlap function was employed

for ensuring that good-quality reads will be considered for the downstream analysis. The sorted and .bam clipOverlapped files were indexed and the REDItoolsDenovo.py script was employed for calling SNVs in the concordant reads. Output data were joined with RNA-seq and MultiEditR data by sample, genomic coordinate, strand, and base identity.

#### Programmable RNA and DNA editing

For the EGFP reactivation experiment, HEK293T cells were transfected with CAGX (50 ng), 4λN-hADAR2-DD-NLS or 4λN-hADAR2-DD (100 ng), and pENTR-gRNA W58X (500 ng). The transfections were performed in a 24-well plate using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. 96 h after transfection, half of the cells were analyzed by fluorescence-activated cell sorting (FACS) to detect the percentage of EGFP-positive cells, and from the other half RNA was extracted and amplified the fragment of EGFP containing the W58X mutation (oligos #18–19). After PCR clean up (Macherey-Nagel NucleoSpin gel and PCR clean-up kit) the fragments were subjected to Sanger sequencing (Eurofins Genomics, GATC services, Germany). A-to-G editing was quantified with MultiEditR with the CAGX alone .ab1 set as the control file. See Figure S7C for analysis parameters.

For T cell base editing, NGS values, flow cytometry values, and .ab1 files were taken from the supplemental information of our previous work.[21] Sequencing .ab1 files were analyzed by MultiEditR using default parameters and analyzed against NGS and flow cytometry values.

#### MultiEditR development

To distinguish edits from background noise, we modified the null hypothesis significance testing (NHST) algorithm from our previous work to operate on multiple edits spread across an amplicon, which we named MultiEditR. MultiEditR requires (1) a sample .ab1 file of an amplicon of interest between approximately 350 and 750 bp, (2) either a control .fasta file of the sequence of the amplicon of interest without any editing events or a control .ab1 of the amplicon of interest without any editing events, (3) a motif of interest consisting of any length of IUPAC nucleotides (e.g., YAR, TCA, A, C, $N_{20}$, $N_n$), (4) a discrete base of interest hypothesized to be edited (e.g., A, C, T, or G), and (5) any hypothesized edited outcomes separated by "|" (e.g., G, T|G, A|T|G).

The MultiEditR algorithm begins by loading the sample .ab1 and trimming the ends of the sequence based on a Phred score cutoff (default of 0.0001) using a modified Mott's algorithm (http://www.phrap.org/phredphrap/phred.html). If a control .ab1 file is used, the file is trimmed in the same manner. Once trimmed, the base calls are extracted from the sample chromatogram and aligned to the non-edited control sequence, where the control sequence index is joined to the sample. Any positions with indels in the alignment are filtered out of the analysis. The motif of interest is then matched to the control sequence, and the control indices where matches are found are used to separate the sample into the alternative sample where matches are found, and the null sample where matches are not found. The noise height of the edited bases of interest (e.g., G if interested in A-to-G edits, or T, G, or A if interested in C-to-T|G|A edits) is extracted from the trace of the null sample. These noise samples are used to model zero-adjusted (zΓ) gamma distributions, which are used as null distributions for the NHST, wherein the p value determines the critical value within the distribution of calling significance versus non-significance, as previously described.[9] The height of hypothetical edits within the motifs of interest (e.g., height of G under A peaks) are then compared to the critical value generated by the null zΓ distributions. If a hypothetical edit is at or above the critical value it will be called as significant and reported as an edit. We made MultiEditR available as web application (https://moriaritylab.shinyapps.io/MultiEditR), and the source code is also available for running the application locally (https://github.com/MoriarityLab/MultiEditR). The application provides diagnostics plots of the sample, visualization, and tabulation of detected edits, a summary of the zΓ modeling, and the ability to download the output analysis data as a tab delimited file (see Figure S2 for visual layout of the web app algorithm).

MultiEditR was written in the statistical programming language R (v3.4.2) using RStudio (v1.1.383). The MultiEditR web app was developed using R shiny (https://shiny.rstudio.com/). global.R contains the definition of functions required to manipulate input files with input parameters and return output analyses, dependencies.R specifies the required packages for the web app, most notably *sangerseqR*[31] for reading and analyzing .ab1 files (https://bioconductor.org/packages/release/bioc/html/sangerseqR.html), *gamlss*[32] for zero adjusted gamma distribution modeling (https://www.gamlss.com/), *tidyverse* packages for data manipulation and visualization (https://www.tidyverse.org/), and *shiny* for support of the web application. server.R interfaces the inputs from the user side to the server side via the functions defined in global.R. ui.R specifies the visual interface of the web app.

#### Data analysis

All statistical analyses were performed in RStudio. The level of significance was set at $\alpha = 0.01$. Student's one-sample, two-tailed t tests were used as indicated in the text. Data were subjected to assumptions of homoscedasticity prior to testing. Data were visualized in RStudio employing various tidyverse (https://www.tidyverse.org/) and Bioconductor (https://www.bioconductor.org/) packages. See https://github.com/MoriarityLab/MultiEditR for reproducible analysis.

#### Availability

The MultiEditR web app is available at z.umn.edu/multieditr. Source code for running the application locally and recreating figures and analyses is available at https://github.com/MoriarityLab/MultiEditR. Original NGS data have been deposited in the NCBI GEO database under accessions GEO: GSE164211 and GSE145011.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2021.07.008.

## ACKNOWLEDGMENTS

We thank the High Throughput Sequencing unit of the Genomics & Proteomics Core Facility, as well as the Flow Cytometry unit of the Imaging and Cytometry Core Facility, German Cancer Research Center (DKFZ), for providing excellent sequencing and sorting services. We also thank Prof. Nina Papavasiliou (DKFZ), Derek Nedveck, and Walker Lahr for helpful conversations surrounding the topic. This work was supported by the Childrens Cancer Research Fund, the Fanconi Anemia Research Foundation, the University of Minnesota Academic Investment Research Program, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 649019 to Prof. Nina Papavasiliou [DKFZ]). This work was made possible by an NIH-funded predoctoral fellowship to Mitchell G. Kluesner (T32GM007266). The graphical abstract was created with BioRender.com.

## AUTHOR CONTRIBUTIONS

M.K., R.P., and B.S.M. designed the experiments. M.B., S.W., and T.L. developed KO cell lines. R.P. and A.A. performed titration experiments and RNA base-editing experiments and sequencing. M.K. wrote the program. R.N.T. defined the pipeline for NGS data processing. M.K., R.N.T., and R.P. analyzed the data. M.K. and R.P. wrote the manuscript. M.K., R.P., and B.S.M. supervised the research. All authors have read and approved the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

1. Eisenberg, E., and Levanon, E.Y. (2018). A-to-I RNA editing—Immune protector and transcriptome diversifier. Nat. Rev. Genet. *19*, 473–490.
2. Lerner, T., Papavasiliou, F.N., and 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.
3. Xu, L.-D., and Öhman, M. (2018). ADAR1 editing and its role in cancer. Genes (Basel) *10*, 12.
4. Reardon, S. (2020). Step aside CRISPR, RNA editing is taking off. Nature *578*, 24–27.
5. Diroma, M.A., Ciaccia, L., Pesole, G., and Picardi, E. (2019). Elucidating the editome: Bioinformatics approaches for RNA editing detection. Brief. Bioinform. *20*, 436–447.
6. Ramaswami, G., and Li, J.B. (2016). Identification of human RNA editing sites: A historical perspective. Methods *107*, 42–47.
7. Toung, J.M., Lahens, N., Hogenesch, J.B., and Grant, G. (2014). Detection theory in identification of RNA-DNA sequence differences using RNA-sequencing. PLoS ONE *9*, e112040.
8. Montiel-González, M.F., Vallecillo-Viejo, I.C., and Rosenthal, J.J.C. (2016). An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res. *44*, e157.
9. Merkle, T., Merz, S., Reautschnig, P., Blaha, A., Li, Q., Vogel, P., Wettengel, J., Li, J.B., and Stafforst, T. (2019). Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. Nat. Biotechnol. *37*, 133–138.
10. Vogel, P., Moschref, M., Li, Q., Merkle, T., Selvasaravanan, K.D., Li, J.B., and Stafforst, T. (2018). Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. Nat. Methods *15*, 535–538.
11. Qu, L., Yi, Z., Zhu, S., Wang, C., Cao, Z., Zhou, Z., Yuan, P., Yu, Y., Tian, F., Liu, Z., et al. (2019). Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. Nat. Biotechnol. *37*, 1059–1069.
12. Rinkevich, F.D., Schweitzer, P.A., and Scott, J.G. (2012). Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. BMC Res. Notes *5*, 63.
13. Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: A method to quantify base editing from Sanger sequencing. CRISPR J. *1*, 239–250.
14. Porath, H.T., Carmi, S., and Levanon, E.Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. Nat. Commun. *5*, 4726.
15. Picardi, E., and Pesole, G. (2013). REDItools: High-throughput RNA editing detection made easy. Bioinformatics *29*, 1813–1814.
16. Picardi, E., D'Erchia, A.M., Gallo, A., and Pesole, G. (2015). Detection of post-transcriptional RNA editing events. Methods Mol. Biol. *1269*, 189–205.
17. Parker, L.T., Zakeri, H., Deng, Q., Spurgeon, S., Kwok, P.Y., and Nickerson, D.A. (1996). AmpliTaq DNA polymerase, FS dye-terminator sequencing: Analysis of peak height patterns. Biotechniques *21*, 694–699.
18. Zakeri, H., Amparo, G., Chen, S.M., Spurgeon, S., and Kwok, P.Y. (1998). Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing. Biotechniques *25*, 406–410, 412–414.
19. Roth, S.H., Levanon, E.Y., and Eisenberg, E. (2019). Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. Nat. Methods *16*, 1131–1138.
20. Vallecillo-Viejo, I.C., Liscovitch-Brauer, N., Montiel-Gonzalez, M.F., Eisenberg, E., and Rosenthal, J.J.C. (2018). Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. RNA Biol. *15*, 104–114.
21. Webber, B.R., Lonetree, C.L., Kluesner, M.G., Johnson, M.J., Pomeroy, E.J., Diers, M.D., Lahr, W.S., Draper, G.M., Slipek, N.J., Smeester, B.A., et al. (2019). Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. Nat. Commun. *10*, 5222.
22. Wang, X., Tilford, C., Neuhaus, I., Mintier, G., Guo, Q., Feder, J.N., and Kirov, S. (2017). CRISPR-DAV: CRISPR NGS data analysis and visualization pipeline. Bioinformatics *33*, 3811–3812.
23. Potapov, V., Fu, X., Dai, N., Corrêa, I.R., Jr., Tanner, N.A., and Ong, J.L. (2018). Base modifications affecting RNA polymerase and reverse transcriptase fidelity. Nucleic Acids Res. *46*, 5753–5763.
24. Severi, F., and Conticello, S.G. (2015). Flow-cytometric visualization of C>U mRNA editing reveals the dynamics of the process in live cells. RNA Biol. *12*, 389–397.
25. Rosenberg, B.R., Hamilton, C.E., Mwangi, M.M., Dewell, S., and Papavasiliou, F.N. (2011). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. Nat. Struct. Mol. Biol. *18*, 230–236.
26. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.
27. Pestal, K., Funk, C.C., Snyder, J.M., Price, N.D., Treuting, P.M., and Stetson, D.B. (2015). Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. Immunity *43*, 933–944.
28. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.
29. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.
30. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
31. Hill, J.T., Demarest, B.L., Bisgrove, B.W., Su, Y.C., Smith, M., and Yost, H.J. (2014). Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. Dev. Dyn. *243*, 1632–1636.
32. Rigby, R.A., Stasinopoulos, D.M., and Lane, P.W. (2005). Generalized additive models for location, scale and shape. J. R. Stat. Soc. Ser. C Appl. Stat. *54*, 507–554.

**OMTN, Volume *25***

## Supplemental information

# MultiEditR: The first tool for the detection and quantification of RNA editing from Sanger sequencing demonstrates comparable fidelity to RNA-seq

**Mitchell G. Kluesner, Rafail Nikolaos Tasakis, Taga Lerner, Annette Arnold, Sandra Wüst, Marco Binder, Beau R. Webber, Branden S. Moriarity, and Riccardo Pecori**

**SUPPLEMENTARY MATERIALS**

**pJET-CmAG-WT and pJET-CmAG-6x**

CTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCC
ACTCCACCGGCGGCATGGACGAGCTGTACAAGaagcttaAACAAGTAGCTGGTGCCAAG
GAAAAAATAACTTCTTTCATGGAAAATTATAGAATTACAGATAATGATGTACTAATT
GCCATAGATAGTGCCAAAATCAACTTCAATGAAAAA**C**TCTCTCAA**C**TTGAGACATA
CGCGATA**C**AATTTGATCAGTATATTAAAGATAATTATGATCCACATGA**C**TTAAAAAG
AA**C**TATTGCTGAGATTATTGAT**C**GAATCATTGAAAAGTTAAAAATTCTTGATGAACA
GTATCATATCCGTGTAAATCTAGCAAAATCAATCCATAATCTCTATTTATTTGTTGAA
AACGTTGATCTTAACCAAGTCAGTAGTAGTAACACCTCTTGGATCCAAAATGTGGAT
TCCAATTATCAAGTCAGAATCCAAATTCAAGAAAAACTACAGCAGCTCAGGACACA
AATTCAGAATATAGACATTCAGCAGCTTGCTGCAGAGGTAAAACGACAG

In **yellow** are highlighted the 6 Cs which are unedited (C) in pJET-CmAG-WT and edited (T) in the pJET-CmAG-6x.

**mCherry-mApob-eGFP (CmAG) coding region:**

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTT
CAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGG
GCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGT
GGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAG
GCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAG
GGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGAC
CCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCA
CCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCC
TCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAG
GCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGG
CCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATC
ACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCG
CCACTCCACCGGCGGCATGGACGAGCTGTACAAGaagcttaAACAAGTAGCTGGTGCCA
AGGAAAAAATAACTTCTTTCATGGAAAATTATAGAATTACAGATAATGATGTACTAA
TTGCCATAGATAGTGCCAAAATCAACTTCAATGAAAAACTCTCTCAACTTGAGACAT
ACGCGATACAATTTGATCAGTATATTAAAGATAATTATGATCCACATGACTTAAAAA
GAACTATTGCTGAGATTATTGATCGAATCATTGAAAAGTTAAAAATTCTTGATGAAC
AGTATCATATCCGTGTAAATCTAGCAAAATCAATCCATAATCTCTATTTATTTGTTGA
AAACGTTGATCTTAACCAAGTCAGTAGTAGTAACACCTCTTGGATCCAAAATGTGGA

TTCCAATTATCAAGTCAGAATCCAAATTCAAGAAAAACTACAGCAGCTCAGGACAC
AAATTCAGAATATAGACATTCAGCAGCTTGCTGCAGAGGTAAAACGACAGACCCGG
GATCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACC
ACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTG
CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCC
ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACC
ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCAC
TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
CAAGTAA

In red mCherry, in yellow mouse Apob and in green eGFP.

**mCherry-Apob-eGFP W58X (CAGX) coding region:**

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTT
CAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGG
GCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGT
GGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAG
GCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAG
GGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGAC
CCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCA
CCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCC
TCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAG
GCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGG
CCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATC
ACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCG
CCACTCCACCGGCGGCATGGACGAGCTGTACAAGAAGCTTACCATGGCCAAGGAGA
AACTGACTGCTCTCACAAAAAAGTATAGAATTACAGAAAATGATATACAAATTGCA
TTAGATGATGCCAAAATCAACTTTAATGAAAAACTATCTCAACTGCAGACATATATG
ATACAATTTGATCAGTATATTAAAGATAGTTATGATTTACATGATTTGAAAATAGCT
ATTGCTAATATTATTGATGAAATCATTGAAAAATTAAAAAGTCTTGATGAGCACTAT
CATATCCGTGTAAATTTAGTAAAAACAATCCATGATCTACATTTGTTTATTGAAAAT
ATTGATTTTAACAAAAGTGGAAGTAGTACTGCATCCTGGATTCAAAATGTGGATACT

AAGTACCAAATCAGAATCCAGATACAAGAAAAACTGCAGCAGCTTAAGAGACACAT
ACAGAATATAGACATCCAGCACCTAGCTGGAATTCTGCAGTCGACGGTACCGCGGG
CCCGGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGT
GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCT
GCACCACCGGCAAGCTGCCCGTGCCCTAGCCCACCCTCGTGACCACCCTGACCTACG
GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGT
CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC
CGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG
CTGTACAAGTAA

In red mCherry, in yellow human Apob and in green eGFP. In light blue is highlighted W58X mutation and in red the A to edit to G to re-activate the eGFP.

## SUPPLEMENTARY TABLES

**Supplementary Table 1 - List of all the regions within transcripts analysed with REDItools v1 for NGS comparison with MultiEditR.** For each transcript strandness, reference genome and chromosomal coordinates of the amplicon are indicated.

| Gene (strand) | Reference Genome | Chromosomal coordinates of the amplicon |
|---|---|---|
| ARSD (reg.1) (-) | hg19 | chrX:2,824,741-2,825,282 |
| ARSD (reg.2) (-) | hg19 | chrX:2,824,114-2,824,494 |
| CTSS (-) | hg19 | chr1:150,704,237-150,704,783 |
| DDX58 (-) | hg19 | chr9:32,455,300-32,502,734 |
| MAVS (+) | hg19 | chr20:3,827,446-3,856,770 |
| SSR3 (-) | hg19 | chr3:156,257,929-156,272,973 |
| Slc39a10 (-) | mm10 | chr1:46,807,710-46,808,199 |
| Aldoc (+) | mm10 | chr11:78,327,206-78,327,625 |
| Atp6ap2 (+) | mm10 | chrX:12,615,885-12,616,656 |
| Serinc1 (-) | mm10 | chr10:57,515,845-57,516,141 |
| B2m (+) | mm10 | chr2:122,147,686-122,153,083 |
| Rab7 (-) | mm10 | chr6:87,999,398-87,999,968 |
| Reep5 (-) | mm10 | chr18:34,346,596-34,347,035 |
| Casp6 (+) | mm10 | chr3:129,913,560-129,914,011 |

**Supplementary Table 2 - List of all the regions within transcripts analysed by Amplicon-seq.** For each transcript strandness, reference genome and chromosomal coordinates of the amplicon are indicated.

| Gene (strand) | Reference Genome | Chromosomal coordinates of the amplicon |
|---|---|---|
| ARSD (-) | hg19 | chrX:2824741-2825282 |
| CTSS (-) | hg19 | chr1:150704237-150704783 |
| DDX58 (-) | hg19 | chr9:32456205-32456648 |
| MAVS (+) | hg19 | chr20:3851441-3851945 |
| Slc39a10 (-) | mm10 | chr1:46807710-46808199 |
| Aldoc (+) | mm10 | chr11:78327206-78327625 |
| Atp6ap2 (+) | mm10 | chrX:12616190-12616656 |
| Serinc1 (-) | mm10 | chr10:57515845-57516325 |
| B2m (+) | mm10 | chr2:122152650-122153068 |
| Rab7 (-) | mm10 | chr6:87999398-87999968 |

**Supplementary Table 3 - List of all the primers and oligos used in this study.** For all the primers and oligos listed above lowercase bases represent the homology part needed during the cloning (#1-8, 17) or the adaptor sequence needed for Amplicon-seq (#57-76). For oligos #13-14 the base in bold is the one modified during the site-directed mutagenesis. In oligo #17 gRNA the bases highlighted in green codify the boxB hairpin. Oligos #27-56 were used to generate the PCR fragments used in the comparison of MultiEditR and NGS methods (Figure 2 and S4).

| | | |
|---|---|---|
| 1 | mApob_F | ggcggcatggacgagctgtacaagAAGCTTAAACAAGTAGCTGGTGCCAAGGAA |
| 2 | mApob_R | accatggtggcgaccggtggatcccgggtCTGTCGTTTTACCTCTGCAGCAAG |
| 3 | RBM47_F | ttagtgaaccgtcagatccgctagcATGACTGCTGAAGATTCCGCC |
| 4 | RBM47_R | agagtcgcggccgctttacttgtacaTCAGTAAGTCTGGTAGACGTCG |
| 5 | CRISPR sgRNA ADAR1ko_ exon4_F | caccGGACAGGAGACGGAATTCGC |
| 6 | CRISPR sgRNA ADAR1ko_ exon4_R | aaacGCGAATTCCGTCTCCTGTCC |
| 7 | non-targeting sgRNA_F | caccGTATTACTGATATTGGT |
| 8 | non-targeting sgRNA_R | aaacACCAATATCAGTAATAC |
| 9 | A1KO_11_fw | CACCGAGCAAGATGAGTTCCGAGAC |

| | | |
|---|---|---|
| 10 | A1KO_11_rv | AAACGTCTCGGAACTCATCTTGCTC |
| 11 | A1KO_39_fw | CACCGTAGCTGTTGATCCCACTCTG |
| 12 | A1KO_39_rv | AAACCAGAGTGGGATCAACAGCTAC |
| 13 | W58X_uAg_F | CTGCCCGTGCCCTAGCCCACCCT |
| 14 | W58X_uAg_R | AGGGTGGGCTAGGGCACGGGCAG |
| 15 | pENTR_plasm_F | TTTTTTCTAGACCCAGCTTTCTTGTA |
| 16 | pENTR_plasm_R | GGTGTTTCGTCCTTTCCACA |
| 17 | Oligo_gRNA_W58X | tggaaaggacgaaacaccTCAGGGTGGTGGCCCTGAAAAAGGGCCGAGGGTGGGCCAGGGCACGGGCAGCTTGCGGCCCTGAAAAAGGGCCTGGTGCAGATttttttctagacccagctt |
| 18 | Apob_F | TGCATCCTGGATTCAAAATGTGG |
| 19 | eGFP_R | TTGAAGTCGATGCCCTTCAG |
| 20 | A1_KO1A_fw | CATTGATGGCTCTGTGGGTGTTC |
| 21 | A1_KO1A_rv | GCTGAAAAGCACCCAGGGAC |
| 22 | A1_KO2A_fw | GTACCTCTCAGATCCTTTGAGAAGTC |

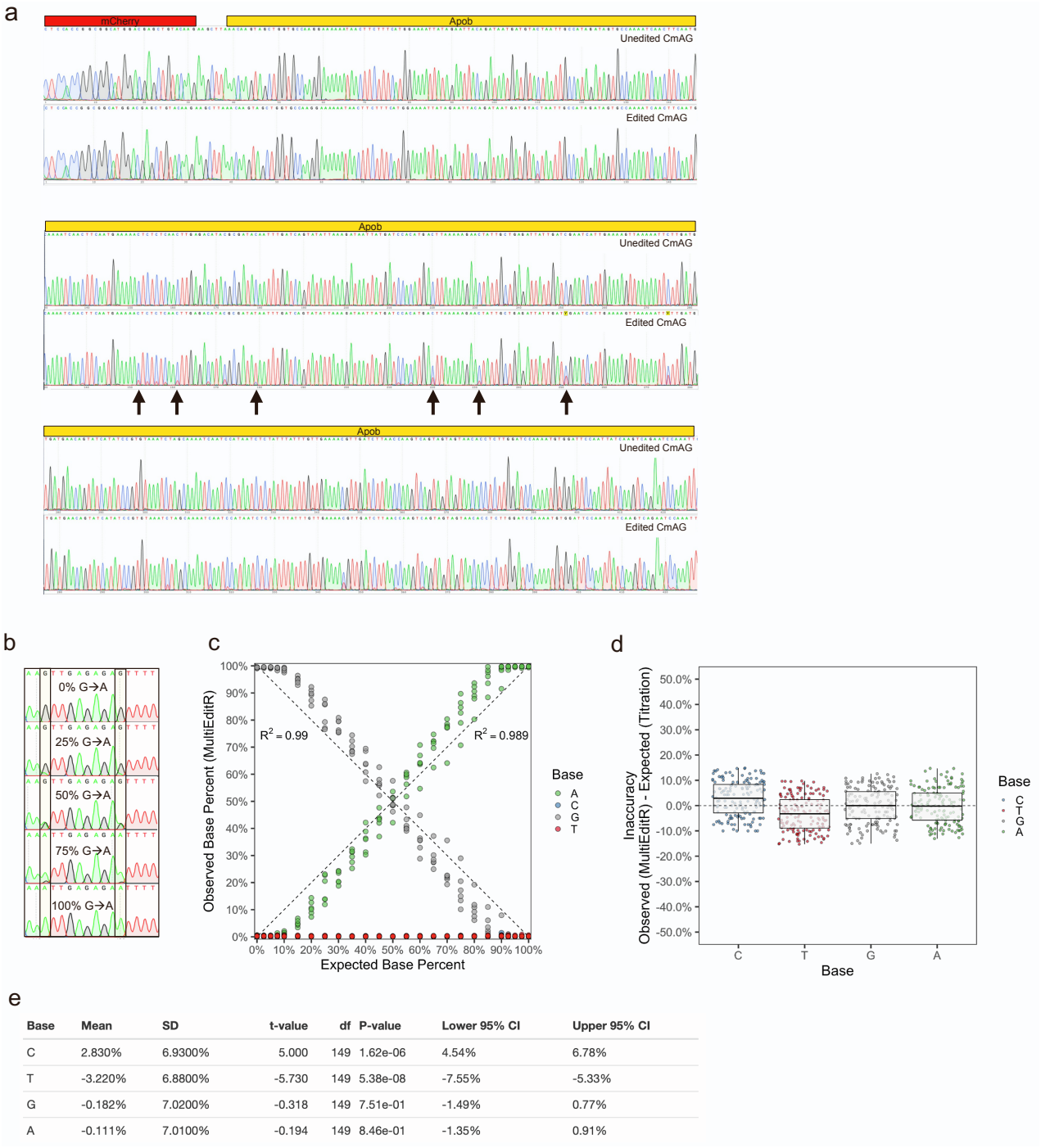
| | | |
|---|---|---|
| 23 | A1_KO2A_rv | GCATGCTGTAACCCTGTAGTTC |
| 24 | B2m_fw | CAAGCATCATGATGCTCTGAAG |
| 25 | B2m_rv | GTAAAAGTAACAAAAGCAGAAGTAGCC |
| 26 | mCherry-mApob-eGFP_SEQ | CTCCCACAACGAGGACTACACC |
| 27 | MAVS_F | TACCCTGCCTGGCCTCAAACTATTA |
| 28 | MAVS_R | ACTTCATGCTGTCTGGGAGCAA |
| 29 | DDX58_F | ATTTGGCCCTGTTGAGCACTCT |
| 30 | DDX58_R | ACGTCCAGGAAACCGCAAACTA |
| 31 | SSR3_F | TGGGCTCCATGCAAGAACTTGGAA |
| 32 | SSR3_R | GCACTGCCATCTAGTGGCAAGTTT |
| 33 | ARSD_reg1_F | GAGCCTGACTGCGTTGCAAACAAA |
| 34 | ARSD_reg1_R | AAGACGGAGGGGTGAAAACATCG |
| 35 | ARSD_reg2_F | CACCCCTCCGTCTTAAAGCATTGT |

| | | |
|---|---|---|
| 36 | ARSD_reg2_R | CCCTTGATCGGCTCATTACTTGGA |
| 37 | CTSS_F | CTGGTGATGGGGTTTAGCAACT |
| 38 | CTSS_R | TGCCAAATAAAGGCCCTGGTCA |
| 39 | Aldoc_F | AAGGGCTATGACCCACTTCCATGT |
| 40 | Aldoc_R | CGATTCCAATTCGAGCGATTGAGG |
| 41 | Atp6ap2_F | GCTAGACTTAGACAACAGGTTTGG |
| 42 | Atp6ap2_R | GTCTACAGATTGAAGCCATACCAC |
| 43 | B2m_F | CAAGCATCATGATGCTCTGAAG |
| 44 | B2m_R | GTAAAAGTAACAAAAGCAGAAGTAGCC |
| 45 | Casp6_F | AACCTAGCAAGTAGGGCCATCTGT |
| 46 | Casp6_R | CATGACCAAGTCAAATAGGCCCAC |
| 47 | Cd36_F | CTGGCTACATCTTTGGTAAAGCCG |
| 48 | Cd36_R | GGGCCACCCCAGTCATGATAG |
| 49 | Rab7_F | TCCGTTCTGAGCAGGCTGTTTTGT |

| | | |
|---|---|---|
| 50 | Rab7_R | TCTGGTGGGTTCTCCTTTCCTCTT |
| 51 | Reep5_F | GCCTTGGAAACTTCCCGCTGTATT |
| 52 | Reep5_R | GTCTCCCTTCTCAGGTCACAGGT |
| 53 | Serinc1_F | TGTATCGCTGCTGTCCAGCAT |
| 54 | Serinc1_R | GGCTGGAACATGAAGATGAACTGC |
| 55 | Slc39a10_F | ACTGCGGGAATATACTCCCACTCT |
| 56 | Slc39a10_R | GCAAGTATCACCTTGCAGGACCAT |
| 57 | ARSD_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagGAGCCTGACTGCGTTGCAAACAAA |
| 58 | ARSD_NGS_R | gtctcgtgggctcggagatgtgtataagagacagAAGACGGAGGGGTGAAAACATCG |
| 59 | CTSS_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagCTGGTGATGGGGTTTAGCAACT |
| 60 | CTSS_NGS_R | gtctcgtgggctcggagatgtgtataagagacagTGCCAAATAAAGGCCCTGGTCA |
| 61 | DDX58_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagATTTGGCCCTGTTGAGCACTCT |
| 62 | DDX58_NGS_R | gtctcgtgggctcggagatgtgtataagagacagGAGAACAATGGCACACGTTAAGAG |
| 63 | MAVS_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagCCATGAGCCCATCACCCAAC |

| | | |
|---|---|---|
| 64 | MAVS_NGS_R | gtctcgtgggctcggagatgtgtataagagacagACTTCATGCTGTCTGGGAGCAA |
| 65 | Slc39a10_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagACTGCGGGAATATACTCCCACTCT |
| 66 | Slc39a10_NGS_R | gtctcgtgggctcggagatgtgtataagagacagGCAAGTATCACCTTGCAGGACCAT |
| 67 | Aldoc_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagAAGGGCTATGACCCACTTCCATGT |
| 68 | Aldoc_NGS_R | gtctcgtgggctcggagatgtgtataagagacagCGATTCCAATTCGAGCGATTGAGG |
| 69 | Atp6ap2_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagGGAGAACGCACTGGGTTTCTTA |
| 70 | Atp6ap2_NGS_R | gtctcgtgggctcggagatgtgtataagagacagGTCTACAGATTGAAGCCATACCAC |
| 71 | Serinc1_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagAGGCTCGGGTTAGGCACTAAGATA |
| 72 | Serinc1_NGS_R | gtctcgtgggctcggagatgtgtataagagacagGGCTGGAACATGAAGATGAACTGC |
| 73 | B2m_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagcaagcatcatgatGCTCTGAAG |
| 74 | B2m_NGS_R | gtctcgtgggctcggagatgtgtataagagacagGTAAAAGTAACAAAAGCAGAAGTAGCC |
| 75 | Rab7_NGS_F | tcgtcggcagcgtcagatgtgtataagagacagTCCGTTCTGAGCAGGCTGTTTTGT |
| 76 | Rab7_NGS_R | gtctcgtgggctcggagatgtgtataagagacagTCTGGTGGGTTCTCCTTTCCTCTT |

## SUPPLEMENTARY FIGURES




e

| Base | Mean | SD | t-value | df | P-value | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|---|---|---|
| C | 2.830% | 6.9300% | 5.000 | 149 | 1.62e-06 | 4.54% | 6.78% |
| T | -3.220% | 6.8800% | -5.730 | 149 | 5.38e-08 | -7.55% | -5.33% |
| G | -0.182% | 7.0200% | -0.318 | 149 | 7.51e-01 | -1.49% | 0.77% |
| A | -0.111% | 7.0100% | -0.194 | 149 | 8.46e-01 | -1.35% | 0.91% |

**Fig. S1 | C-to-U editing of Apob by APOBEC1 within CmAG plasmid.**

**a**, Sanger sequencing chromatograms from HEK293T cells transfected only with CmAG or with CmAG, APOBEC1 and RBM47. Amplification of Apob region from CmAG transcript and

following Sanger sequencing show abundant C-to-T editing in several sites along the sequenced area. Black arrows point to the 6 Cs which are unedited (C) in pJET-CmAG-WT and edited (T) in the pJET-CmAG-6x (Supplementary materials and Fig. 1a). **b**, Chromatograms from G-to-A titration showing a change in peak height at two sites. **c**, Titration of pJET-CmAG-WT with pJET-CmAG-6x sequenced with the forward primer, coefficient of determination was calculated relative to the identity line y = x, N = 6 sites per chromatogram. **d**, Inaccuracy of MultiEditR relative to expected titration values, significance was determined using Student's one-sample t-test relative to an inaccuracy of 0%. **e**, Table of results from Student's one-sample t-test of inaccuracies.

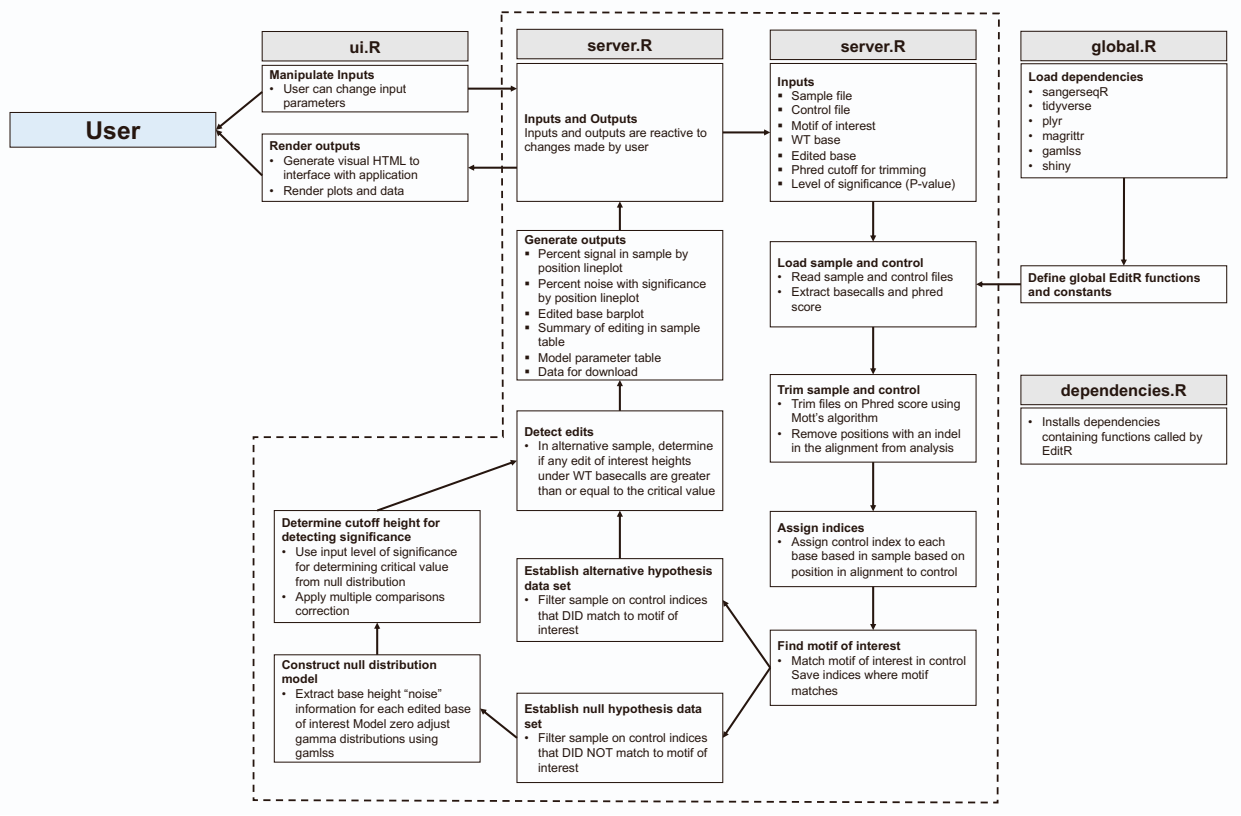



**Fig. S2 | MultiEditR algorithm.**

MultiEditR algorithm flow chart in the context of the MultiEditR web application made using R Shiny (z.umn.edu/multieditr). Source code available at:

https://github.com/MoriarityLab/MultiEditR

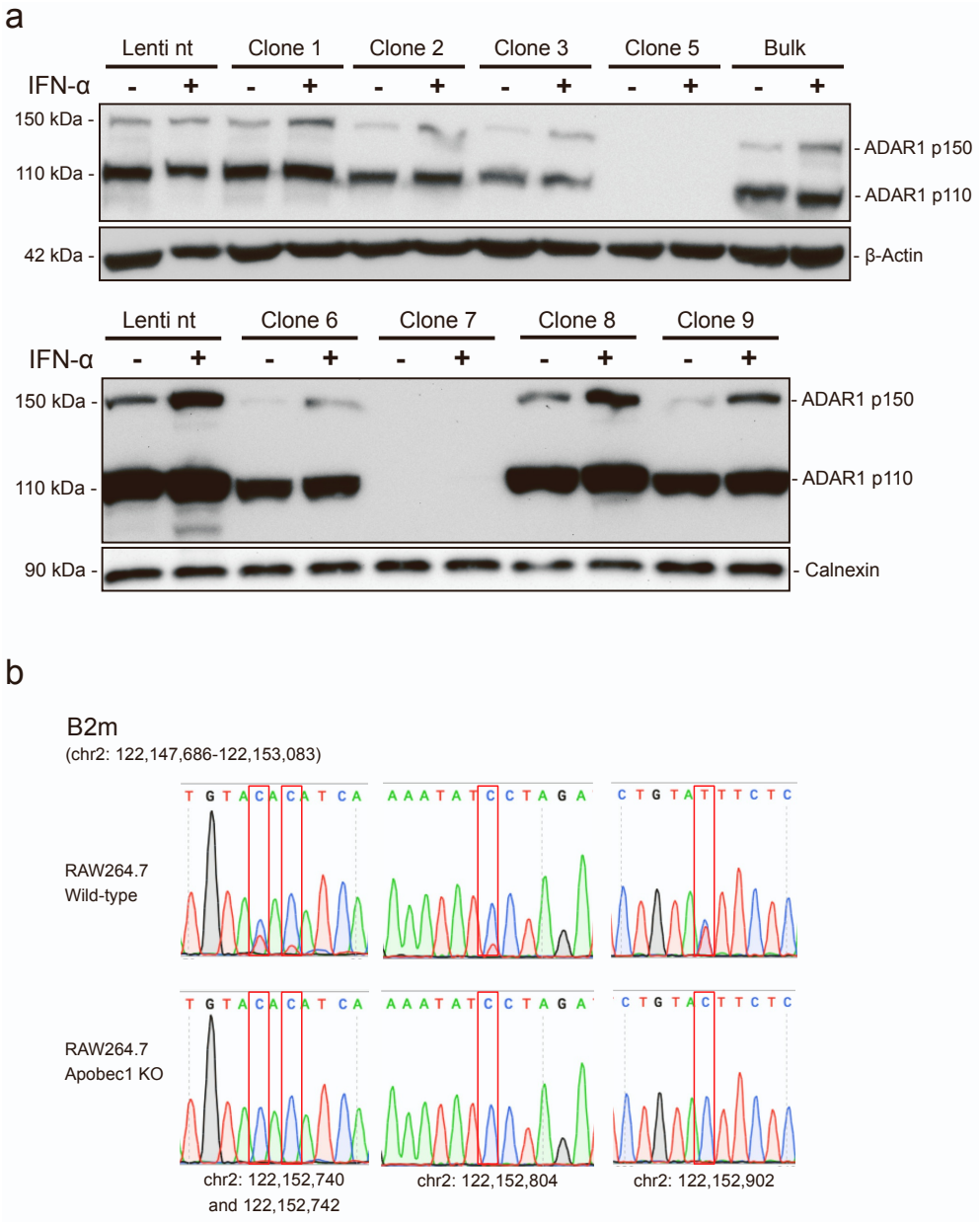
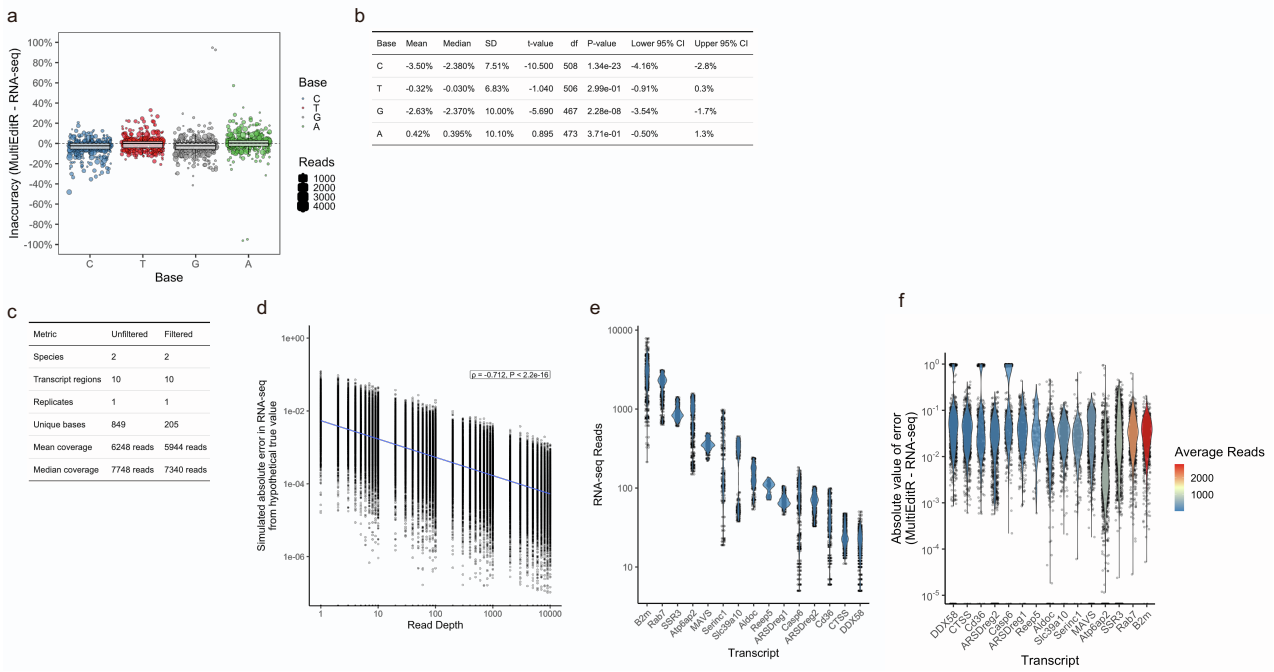



**Fig. S3 | Generation of human ADAR1 and murine APOBEC1 KO cell lines.**

**a**, Western blot of A549 KO clones and bulk, after transduction with lenti-CRISPR-ADAR1 exon 4. A549 cells were stimulated with 200 U/ml of IFN-α (pbl assay science, cat. no.#11100-1) for 16 h to evaluate the presence of ADAR1 p150. "Lenti nt" is the non-targeting control and was generated by transducing cells with the same vector (LentiCRISPRv2) but containing a non-targeting sgRNA (see Supplementary Table 3). β-Actin and Calnexin were used as loading controls

(Mouse β-Actin monoclonal antibody, Sigma-Aldrich, cat. no.#A5441; Rabbit Calnexin polyclonal antibody, Enzo Life science, cat. no.#ADI-SPA-865-F). **b**, Sanger sequencing chromatograms for RAW 264.7 wild-type and APOBEC1 knock-out after RT-PCR amplification of B2m 3'UTR region (oligos #24-25). The complete absence of C-to-U editing along the sequenced region confirms the deficiency of APOBEC1 editing activity within the knock-out clone.



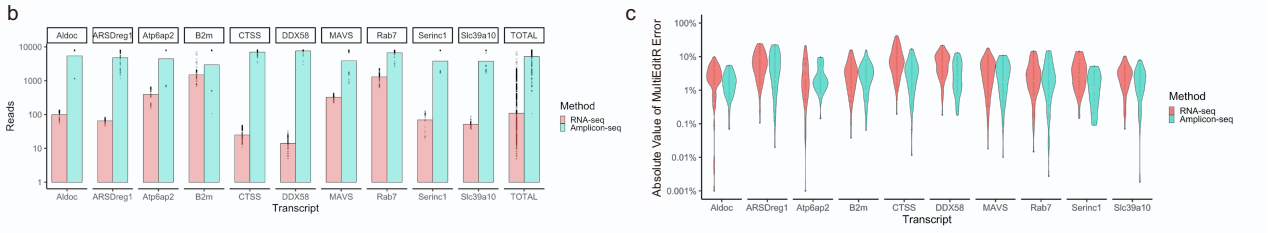

| Base | Mean | Median | SD | t-value | df | P-value | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|---|---|---|---|
| C | -3.50% | -2.380% | 7.51% | -10.500 | 508 | 1.34e-23 | -4.16% | -2.8% |
| T | -0.32% | -0.030% | 6.83% | -1.040 | 506 | 2.99e-01 | -0.91% | 0.3% |
| G | -2.63% | -2.370% | 10.00% | -5.690 | 467 | 2.28e-08 | -3.54% | -1.7% |
| A | 0.42% | 0.395% | 10.10% | 0.895 | 473 | 3.71e-01 | -0.50% | 1.3% |

**Fig. S4 | Descriptive statistics, coverage and error of samples in RNA-seq data set.**

**a-b**, Comparison of MultiEditR to RNA-seq measurements. All data is filtered on between 1% and 99% editing measured by either method. Coefficients of determination represent regression to the identity line. Dot size is proportional to read coverage at the base of interest in RNA-seq. Tables represent results of Student's pairwise t-test between measurement methods. **c**, Descriptive statistics of samples in RNA-seq data set. **d**, Simulation of expected absolute error in RNA-seq editing value from true editing value as a function of read depth. Absolute error of editing measurement is negatively correlated with read depth (Spearman's correlation coefficient, $\rho = -0.712$, $P < 2.2e\text{-}16$). **e**, Violin plots of read depth at each base for each transcript analyzed. **f**, Violin plots of absolute value of error in MultiEditR editing measurement minus REDItools RNA-seq editing measurement by each transcript. Plots are colored by average read depth.

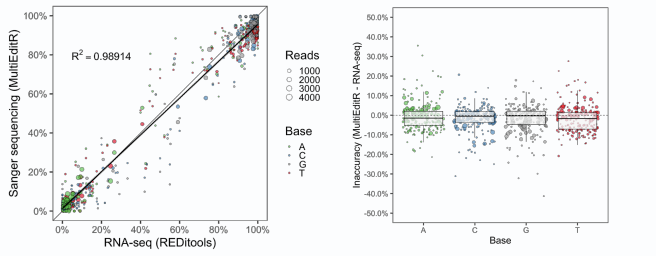
a

| Metric | Unfiltered | Filtered |
|---|---|---|
| Species | 2 | 2 |
| Transcript regions | 10 | 10 |
| Replicates | 1 | 1 |
| Unique bases | 849 | 205 |
| Mean coverage | 6248 reads | 5944 reads |
| Median coverage | 7748 reads | 7340 reads |




d

| Transcript | P-value | Mean difference | Mean RNA-seq Reads | SD RNA-seq Reads | Mean Amplicon-seq Reads | SD Amplicon-seq Reads |
|---|---|---|---|---|---|---|
| Aldoc | 1.60e-06 | 0.922% | 103.0 | 25.40 | 6610 | 2750 |
| ARSDreg1 | 8.29e-02 | 0.592% | 65.6 | 7.46 | 5700 | 2480 |
| Atp6ap2 | 1.61e-02 | 0.828% | 437.0 | 171.00 | 6220 | 3120 |
| B2m | 1.32e-01 | 0.291% | 1710.0 | 1060.00 | 5360 | 3590 |
| CTSS | 0.00e+00 | 5.480% | 26.9 | 10.40 | 7110 | 1060 |
| DDX58 | 0.00e+00 | 3.340% | 15.5 | 6.31 | 7670 | 739 |
| MAVS | 3.97e-04 | 0.566% | 331.0 | 48.10 | 5160 | 3070 |
| Rab7 | 6.02e-04 | 0.244% | 1380.0 | 474.00 | 6900 | 1290 |
| Serinc1 | 5.00e-07 | 2.740% | 80.2 | 34.30 | 4880 | 3060 |
| Slc39a10 | 2.52e-05 | 0.785% | 52.5 | 9.35 | 4610 | 2780 |




g

| Base | Mean | Median | SD | t-value | df | P-value | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|---|---|---|---|
| C | -1.79% | -0.49% | 6.57% | -3.90 | 205 | 1.28e-04 | -2.69% | -0.885% |
| T | -1.92% | -1.72% | 6.89% | -4.06 | 211 | 7.01e-05 | -2.85% | -0.986% |
| G | -2.08% | -0.31% | 7.69% | -3.86 | 202 | 1.54e-04 | -3.15% | -1.020% |
| A | -0.73% | -1.48% | 7.46% | -1.41 | 209 | 1.59e-01 | -1.74% | 0.287% |

h

| Base | Mean | Median | SD | t-value | df | P-value | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|---|---|---|---|
| C | -2.08% | -1.280% | 4.350% | -6.850 | 205 | 8.37e-11 | -2.67% | -1.48% |
| T | -1.66% | -1.270% | 4.720% | -5.110 | 211 | 7.08e-07 | -2.30% | -1.02% |
| G | -2.59% | -1.740% | 5.110% | -7.230 | 202 | 9.96e-12 | -3.30% | -1.89% |
| A | -0.25% | -0.831% | 5.070% | -0.715 | 209 | 4.76e-01 | -0.94% | 0.44% |

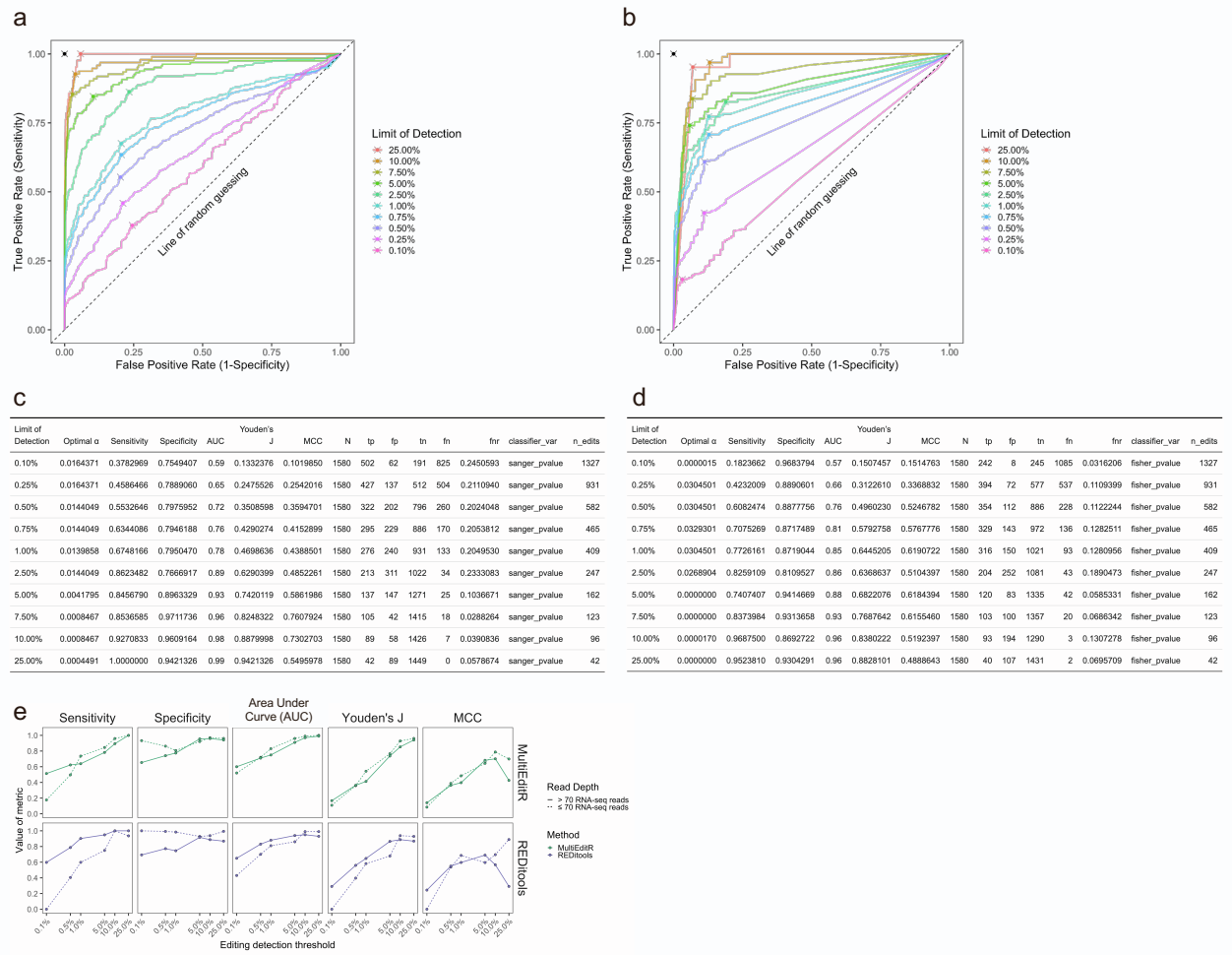
i

| Base | Mean | Median | SD | t-value | df | P-value | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|---|---|---|---|
| C | -0.29% | -0.32% | 5.370% | -0.774 | 205 | 4.40e-01 | -1.03% | 0.45% |
| T | 0.26% | 0.50% | 5.410% | 0.700 | 211 | 4.85e-01 | -0.47% | 0.99% |
| G | -0.51% | -0.56% | 5.550% | -1.310 | 202 | 1.92e-01 | -1.28% | 0.26% |
| A | 0.48% | 0.60% | 5.570% | 1.240 | 209 | 2.15e-01 | -0.28% | 1.24% |

**Fig. S5 | Descriptive statistics, coverage and error of samples in the combined RNA-seq and Amplicon-seq data set.**

**a**, Descriptive statistics of samples in the combined data set. **b**, Barplots of read depth in each sample for RNA-seq and Amplicon-seq data for each transcript. Amplicon-seq read depth is higher for every transcript. **c**, Violin plots of absolute value of error in MultiEditR editing measurements minus REDItools RNA-seq or Amplicon-seq editing measurements. **d**, Table comparison of RNA-seq to Amplicon-seq editing values for each transcript. **e-f**, Comparison of MultiEditR to RNA-seq measurements. Data is filtered on between 1% and 99% editing measured by Amplicon-seq. Coefficients of determination represent regression to the identity line. Dot size is proportional to read coverage at the base of interest in RNA-seq. **g**, Tables of Student's pairwise t-test results

between MultiEditR and RNA-seq. **h**, Tables of Student's pairwise t-test results between MultiEditR and Amplicon-seq. **i**, Tables of Student's pairwise t-test results between RNA-seq and Amplicon-seq.




c

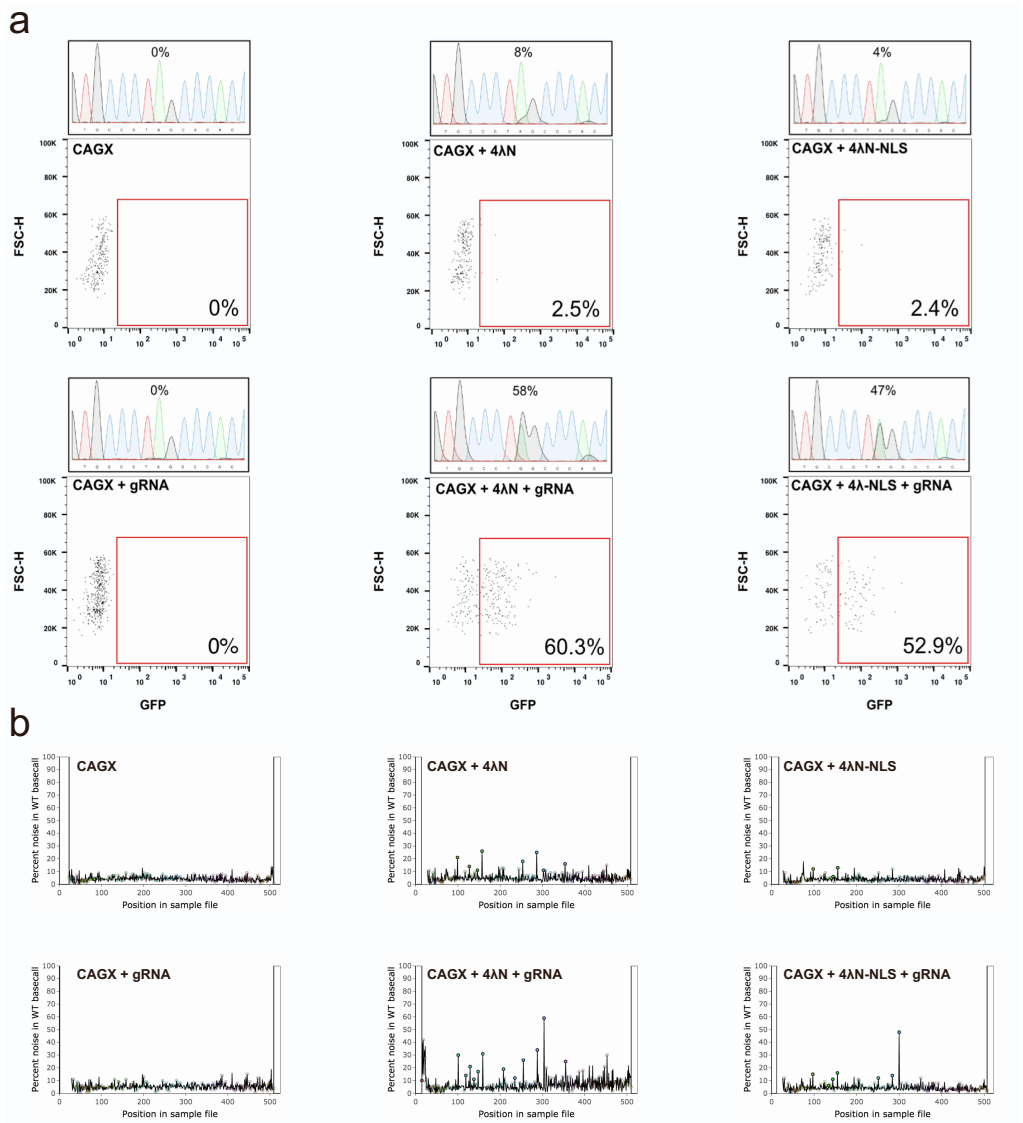
| Limit of Detection | Optimal α | Sensitivity | Specificity | AUC | Youden's J | MCC | N | tp | fp | tn | fn | fnr | classifier_var | n_edits |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.10% | 0.0164371 | 0.3782969 | 0.7549407 | 0.59 | 0.1332376 | 0.1019850 | 1580 | 502 | 62 | 191 | 825 | 0.2450593 | sanger_pvalue | 1327 |
| 0.25% | 0.0164371 | 0.4586466 | 0.7889060 | 0.65 | 0.2475526 | 0.2542016 | 1580 | 427 | 137 | 512 | 504 | 0.2110940 | sanger_pvalue | 931 |
| 0.50% | 0.0144049 | 0.5532646 | 0.7975952 | 0.72 | 0.3508598 | 0.3594701 | 1580 | 322 | 202 | 796 | 260 | 0.2024048 | sanger_pvalue | 582 |
| 0.75% | 0.0144049 | 0.6344086 | 0.7946188 | 0.76 | 0.4290274 | 0.4152899 | 1580 | 295 | 229 | 886 | 170 | 0.2053812 | sanger_pvalue | 465 |
| 1.00% | 0.0139858 | 0.6748166 | 0.7950470 | 0.78 | 0.4698636 | 0.4388501 | 1580 | 276 | 240 | 931 | 133 | 0.2049530 | sanger_pvalue | 409 |
| 2.50% | 0.0144049 | 0.8623482 | 0.7666917 | 0.89 | 0.6290399 | 0.4852261 | 1580 | 213 | 311 | 1022 | 34 | 0.2333083 | sanger_pvalue | 247 |
| 5.00% | 0.0041795 | 0.8456790 | 0.8963329 | 0.93 | 0.7420119 | 0.5861986 | 1580 | 137 | 147 | 1271 | 25 | 0.1036671 | sanger_pvalue | 162 |
| 7.50% | 0.0008467 | 0.8536585 | 0.9711736 | 0.96 | 0.8248322 | 0.7607924 | 1580 | 105 | 42 | 1415 | 18 | 0.0288264 | sanger_pvalue | 123 |
| 10.00% | 0.0008467 | 0.9270833 | 0.9609164 | 0.98 | 0.8879998 | 0.7302703 | 1580 | 89 | 58 | 1426 | 7 | 0.0390836 | sanger_pvalue | 96 |
| 25.00% | 0.0004491 | 1.0000000 | 0.9421326 | 0.99 | 0.9421326 | 0.5495978 | 1580 | 42 | 89 | 1449 | 0 | 0.0578674 | sanger_pvalue | 42 |

d

| Limit of Detection | Optimal α | Sensitivity | Specificity | AUC | Youden's J | MCC | N | tp | fp | tn | fn | fnr | classifier_var | n_edits |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.10% | 0.0000015 | 0.1823662 | 0.9683794 | 0.57 | 0.1507457 | 0.1514763 | 1580 | 242 | 8 | 245 | 1085 | 0.0316206 | fisher_pvalue | 1327 |
| 0.25% | 0.0304501 | 0.4232009 | 0.8890601 | 0.66 | 0.3122610 | 0.3368832 | 1580 | 394 | 72 | 577 | 537 | 0.1109399 | fisher_pvalue | 931 |
| 0.50% | 0.0304501 | 0.6082474 | 0.8877756 | 0.76 | 0.4960230 | 0.5246782 | 1580 | 354 | 112 | 886 | 228 | 0.1122244 | fisher_pvalue | 582 |
| 0.75% | 0.0329301 | 0.7075269 | 0.8717489 | 0.81 | 0.5792758 | 0.5767776 | 1580 | 329 | 143 | 972 | 136 | 0.1282511 | fisher_pvalue | 465 |
| 1.00% | 0.0304501 | 0.7726161 | 0.8719044 | 0.85 | 0.6445205 | 0.6190722 | 1580 | 316 | 150 | 1021 | 93 | 0.1280956 | fisher_pvalue | 409 |
| 2.50% | 0.0268904 | 0.8259109 | 0.8109527 | 0.86 | 0.6368637 | 0.5104397 | 1580 | 204 | 252 | 1081 | 43 | 0.1890473 | fisher_pvalue | 247 |
| 5.00% | 0.0000000 | 0.7407407 | 0.9414669 | 0.88 | 0.6822076 | 0.6184394 | 1580 | 120 | 83 | 1335 | 42 | 0.0585331 | fisher_pvalue | 162 |
| 7.50% | 0.0000000 | 0.8373984 | 0.9313658 | 0.93 | 0.7687642 | 0.6155460 | 1580 | 103 | 100 | 1357 | 20 | 0.0686342 | fisher_pvalue | 123 |
| 10.00% | 0.0000170 | 0.9687500 | 0.8692722 | 0.96 | 0.8380222 | 0.5192397 | 1580 | 93 | 194 | 1290 | 3 | 0.1307278 | fisher_pvalue | 96 |
| 25.00% | 0.0000000 | 0.9523810 | 0.9304291 | 0.96 | 0.8828101 | 0.4888643 | 1580 | 40 | 107 | 1431 | 2 | 0.0695709 | fisher_pvalue | 42 |

**Fig. S6 | Receiver Operating Characteristics (ROC) analysis of MultiEditR and REDItools RNA-seq, compared to REDItools Amplicon-seq.**

**a**, MultiEditR ROC across different limits of detection. Hashed points represent optimal cutpoint. **b**, REDItools RNA-seq ROC across different limits of detection. Hashed points represent optimal cutpoint. **c**, MultiEditR ROC analysis parameter table across levels of detection. **d**, REDItools RNA-seq ROC analysis parameter table across levels of detection. **e**, Lineplots for ROC parameters for MultiEditR and REDItools across limits of detection, broken up by regions that had high RNA seq coverage (>70 reads), and low RNA-seq coverage (≤ 70 reads). MultiEditR is more robust in regions that have poor coverage in RNA-seq.




c

| enzyme | guide | target_edit | gfp | mei | n_sig_edits | ave_sig_edits | ave_nontarget_edits | phred | p_value | target_MEI_ratio | target_nontarget_ratio | norm_target_MEI_ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CAGX | - gRNA | 0.01 | 0.000 | 0.0151 | 0 | 0.0000 | 0.0000 | 1e-04 | 1e-04 | 0.662 | 0.000 | 1.00 |
| CAGX + 4λN | - gRNA | 0.08 | 0.025 | 0.0317 | 5 | 0.1720 | 0.1560 | 1e-03 | 1e-04 | 2.520 | 0.514 | 3.81 |
| CAGX + 4λN-NLS | - gRNA | 0.04 | 0.024 | 0.0179 | 3 | 0.0775 | 0.0642 | 1e-04 | 1e-04 | 2.230 | 0.623 | 3.37 |
| CAGX | + gRNA | 0.03 | 0.000 | 0.0309 | 0 | 0.0000 | 0.0000 | 1e-04 | 1e-04 | 0.971 | 0.000 | 1.47 |
| CAGX + 4λN | + gRNA | 0.58 | 0.603 | 0.0484 | 8 | 0.2570 | 0.1740 | 3e-03 | 1e-04 | 12.000 | 3.330 | 18.10 |
| CAGX + 4λN-NLS | + gRNA | 0.47 | 0.529 | 0.0260 | 4 | 0.1990 | 0.0819 | 3e-03 | 1e-04 | 18.100 | 5.740 | 27.30 |

**Fig. S7 | Reactivation of eGFP by RNA targeted editing.**

**a**, FACS analysis for GFP positive cells after transfection with 4λN and 4λN-NLS in presence and absence of the gRNA to re-activate eGFP by RNA base-editing. The percentage in the plot represents the GFP+ cells of the mCherry+ cells detected by FACS. Above each FACS plot, it is shown A-to-G editing at position W58X for that sample; the percentage of editing is calculated by MultiEditR. **b**, Plots generated in MultiEditR showing the distribution of A-to-G editing along the sequenced region. From this plot appears evident that 4λN has a higher editing activity which leads to a more prominent RNA off-target effect both in presence and absence of the gRNA. **c**, MultiEditR parameters used in this analysis.