MANUSCRIPT TITLE

Developmental editing control of mRNA ND7 involves KREH2 helicase-dependent gRNA selection and proposed RNA structure in *Trypanosoma brucei*

AUTHORS

Joshua Meehan[1]†, Alasdair Ivens[4], Scott Grote,[5], Tyler Rodshagen[2], Zihao Chen[4], Cody Goode[1], Sunil K. Sharma[1], Vikas Kumar[1], Addison Frese[1], Zachary Goodall[1], Laura McCleskey[1], Rebecca Sechrist[1], Lanying Zeng[1], Nicholas J. Savill[4], Silvi Rouskin[5], Achim Schnaufer[4], Suzanne M. McDermott[2,3], Jorge Cruz-Reyes[1*],

[1] Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, 77843, USA.
[2] Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, 98109, USA.
[3] Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, USA.
[4] Institute of Immunology and Infection Research, University of Edinburgh, Edinburgh, EH9 3FL, UK
[5] Department of Microbiology, Harvard Medical School, Boston, MA, 02115, USA.
† Present address: Department of Biochemistry, Therapeutics Innovation Center, Baylor College of Medicine, Houston, TX, 77030, USA.

* Correspondence should be addressed to:

Jorge Cruz-Reyes
cruzrey@tamu.edu
+1-979-458-3374

† Present address: Department of Biochemistry, Therapeutics Innovation Center, Baylor College of Medicine, Houston, TX, 77030, USA.

ABSTRACT

Unknown factors regulate mitochondrial U-insertion/deletion (U-indel) RNA editing in procyclic-form (PCF) and bloodstream-form (BSF) trypanosomes. This editing is directed by anti-sense gRNAs, creates canonical protein-encoding mRNAs, and may developmentally control respiration. Canonical editing by housekeeping cognate gRNAs occurs amid massive non-canonical editing of unclear sources and biological significance. We found PCF-specific repression in mRNA ND7 that involves novel modulation of gRNA usage by RNA helicase KREH2. At a major early checkpoint near the mRNA 3′ terminus, KREH2 promotes the usage of a non-cognate "terminator" gRNA over a cognate gRNA. At this early checkpoint, terminator-directed editing installs a structural element in 30% of the ND7 transcriptome, which potentially occludes the mRNA 3′ terminus and represses further editing. BSF-to-PCF differentiation *in vitro* recreated this form of negative control. Remarkably, KREH2-RNAi knockdown relieved repression by reverting the native usage of cognate/non-cognate gRNAs at the major early checkpoint. ND7 transcripts that lacked terminator-directed editing at the early checkpoint exhibited similar negative editing control along the mRNA sequence, suggesting that KREH2 modulates guiding fidelity globally. Thus, KREH2 is a key protein in ND7 developmental editing control involving modulation of gRNA selection and RNA structure installed by a regulatory terminator gRNA.

INTRODUCTION

Trypanosoma brucei, a member of the protist group Euglenozoa, causes Human African Trypanosomiasis (HAT) (1-3). Its complex life cycle includes procyclic forms (PCF) and bloodstream forms (BSF), which infect insect and mammalian hosts, respectively. Trypanosomatids, including *T. brucei*, exhibit massive post-transcriptional site-specific uridine insertion and deletion (U-indel) RNA editing in their large mitochondrion. Twelve of the 18 primary mRNA transcripts in mitochondria (mt) lack the correct open reading frame (ORF), which is created via U-indel RNA editing directed by trans-acting cognate guide RNAs (gRNAs), except for one case where a gRNA is provided in cis (4,5). The editing process is developmentally regulated; however, the key regulatory factors have not been defined.

Pre-mRNA and gRNA transcripts are encoded in the mt-genomes of trypanosomatids (aka kinetoplast, kDNA), a planar network of catenated relaxed circles (6,7). Maxicircles encode rRNAs, and pre-mRNAs for ribosomal protein S12 (RPS12), ATPase subunit 6 (A6), and proteins in respiratory complexes. Minicircles encode hundreds of cognate gRNAs (~45-60nt) that partially hybridize with pre-mRNA but fully complement the canonical edited sequence of mature mRNAs. A short anchor region in gRNAs initiates binding with mRNA through Watson-Crick pairing, while a guiding region directs the U-indels via Watson-Crick and GU wobble pairing (8,9). Often, gRNAs contain encoded 5' and 3' terminal bases, not used in canonical editing, and a post-transcriptionally added 3' oligo(U) tail. Editing progresses 3' to 5' in overlapping blocks, and each block is directed by a specific gRNA, but typically, there is redundancy, with multiple gRNA species covering the same block (8,10). RT-PCR of the 5'-most block typically examines the accumulation of fully edited mRNA, while detailed editing progression studies, including at early 3' blocks, require nucleotide-resolution RNA-seq (11-13). Most mt-mRNAs are edited extensively, while others are edited moderately or are never edited (14). However, only a few mRNA molecules match the canonical pattern at steady-state, while the vast majority carry unexpected non-canonical edits of unclear function or origin. Non-canonical U-indels are usually found in editing junctions between 3' canonical and 5' pre-edited sequences (15,16) (**Supplementary Table ST1**, glossary of terms).

T. brucei uses dramatically different strategies in ATP production during development (17-21). PCF cells employ cytochrome-mediated oxidative phosphorylation. However, BSF cells employ glycolysis as they lack cytochromes and some Krebs cycle enzymes, and sugar is plentiful in serum. Accordingly, cytochrome-encoding mRNAs (complexes III and IV) with the canonical edited sequence accumulate in PCF but are rarely detectable in BSF. Other edited mRNAs, e.g., for complex I (NADH dehydrogenase), exhibit large differences in abundance between the two stages. Some transcripts, e.g., RPS12 or A6 in the F_1F_O-ATPase complex (complex V), are efficiently edited during all stages of development. This switch in bioenergetics may involve the regulation of editing rather than substrate availability since pre-mRNA and gRNA transcripts are relatively constant at steady-state (5,10,22).

Most of the about 40 known proteins in the editosome holoenzyme are arranged in a few ribonucleoprotein (RNP) complexes and their variants (4,15). Three RNA Editing Catalytic Complexes (RECCs) catalyze endonucleolytic cleavage, U indels, and ligation (23-26). RNA Editing Substrate Binding Complexes (RESCs) (27-30) serve as platforms for mRNA-gRNA hybrid formation and processing (31-35). We and others initially showed that purified RESC is enriched with editing substrates and products (31,33). Two RESC cryoEM reconstructed structures (RESC-A and RESC-B) may represent gRNA storage and mRNA-gRNA duplex substrate complexes, respectively (36,37). RNA Editing Helicase KREH2-associated Complex (REH2C) (4,5,15) contains three core proteins: the typifying DEAH-Box RNA helicase KREH2 and two directly bound helicase factors, the zinc finger protein KH2F1 and KH2F2. KREH2 controls editing fidelity (i.e., the ratio of non-canonical /canonical edits), including the first example of differential gRNA-directed non-canonical editing in PCF and BSF (38) Sedimentation analyses suggested that REH2C may also have different variants (39). Multiple other factors that are not stably associated with the major editing complexes are also required for complete editing. These include a second RNA editing helicase, DEAD-box KREH1, which appears to promote initiator gRNA utilization (40).

Several studies indicate that REH2C is an RNP, and its association with other mt factors could involve KREH2-dependent holo-editosome remodeling. Isolated KREH2 co-purifies with proteins in RESC, RECC, KREH1, and other mt factors via RNA (31,32,41,42). All three core proteins in REH2C co-purified with mRNA substrates and products (fully and partially edited) in the absence of RESC, i.e., upon depletion of RESC1 and gRNA (32). Conversely, several purifications of RESC proteins have detected KREH2 (27-30), supporting the speculation that KREH2 could mediate remodeling between RESC-A and RESC-B CryoEM reconstructions (36). KREH2 immunoprecipitated from mt-extract or recombinantly expressed, crosslinked with RNA, and supported dsRNA unwinding *in vitro* (41,42). Inactivating point mutation in the KREH2 helicase core and dsRNA binding (dsRBD) motifs dissociated KREH2 from RESC, suggesting that a functional helicase mediates REH2C-RESC interaction (32,41,42). Furthermore, KREH2 RNAi-knockdown decreased the accumulation of all fully-edited mature mRNAs examined by RT-qPCR in PCF and BSF (13,32,39,42), decreased editing fidelity in RESC-bound A6 and RPS12 in deep sequencing studies (13,38), and shifted RESC proteins in a sedimentation analysis (43). A study in PCF and BSF lifecycle stages revealed PCF-specific

KREH2 upregulation of a noncognate gRNA usage, which introduces a non-canonical 3′ High-Frequency Element (3′ HFE) in >30% of the A6 transcriptome. This HFE at the A6 3′ terminus blocked all canonical editing in the targeted molecules (38). The latter observation prompted us to test the hypothesis that KREH2 plays a role in modulating the editing of substrates that are known to be developmentally regulated.

In this study, we specifically asked whether KREH2 controls the maturation of mRNA ND7 (NADH dehydrogenase subunit 7). ND7 has two editing domains, of which the 3′-domain (ND7 3′) exhibits preferential maturation in BSF cells by unknown mechanisms (14,44,45), including upon induced *in vitro* differentiation from PCF into mammalian infective forms; however, differences in RT-PCR assays were also reported between strains (46). Here, we carried out nucleotide-resolution RNA-seq, revealing a PCF-specific negative control checkpoint in early ND7 3′ editing in our lab strains. At this early checkpoint, the presence of KREH2 promotes preferential usage of a novel "terminator" gRNA (guide) over the cognate guide at the corresponding block within PCF cells. This terminator installs a non-canonical 3′ HFE in ~30% of the ND7 transcriptome, derailing early editing at the 3′ domain in PCF cells. *In vitro* DMS-MaPseq analyses suggested the 3′ HFE forms a stable RNA fold, likely blocking canonical guide hybridization with the ND7 mRNA. Remarkably, KREH2 knockdown reverted the natural gRNA selection at this checkpoint, i.e., decreased the 3′ HFE and increased canonical editing at the corresponding block, particularly within native RESC6-purified RESC complexes, suggesting the involvement of active KREH2-regulated editosomes. An increase in ND7 canonical editing upon KREH2 loss-of-function challenges the original definition of an editing factor. The KREH2 knockdown also induced similar converse changes between canonical and non-canonical edits along the length of ND7 transcripts that did not receive the 3′ HFE in PCF cells. Thus, KREH2 is the first protein factor shown to have a direct role in developmental editing control. KREH2 may globally control editing fidelity by modulating gRNA selection, i.e., promoting one type of gRNA over another (cognate vs. non-cognate) during the trypanosome lifecycle. Regulated gRNA-programmed non-canonical editing challenges the view of editing as an example of constructive neutral evolution (38). These studies provide insights to understand better key factors and associated mechanisms of developmental U-indel editing control, which includes differential KREH2-dependent gRNA selection.

MATERIAL AND METHODS

General PCF and BSF cell culture and transfection

T. brucei cell lines in this study, strains Lister 427 29-13 PCF and Lister 427 Single Marker BSF (monomorphic), each expressing a tetracycline-inducible KREH2 RNAi-plasmid construct, and Lister 427 29-13 PCF expressing a tetracycline-inducible KH2F1 RNAi-plasmid construct were grown as in our recent study (38). Briefly, PCF and BSF cells were maintained in selective SDM-79 medium and HMI-9 medium, respectively, supplemented with 10% tetracycline (Tet)-tested fetal bovine serum (FBS) (R&D Systems). PCF and BSF cells were maintained in log phase growth with regular dilutions to maintain cell densities of $<1x10^7$ and $<1x10^6$ cells/mL, respectively. Cell lines were selected with G418 Sulfate (Thermo Fisher), Hygromycin (Invivogen), and Phleomycin (Thermo Fisher), as reported (38). RNAi constructs induced with 1 mg/mL Tet (Sigma) for three days (BSF) or three to four days (PCF) downregulated the targeted protein by ~80% in both cell stages, as reported (32,38). Growth curves and Western blotting of editing proteins determined these time points (13,38). Growth curves were carried out in 24-well plates and counted on a Beckman-Coulter Counter Z2. Log phase PCF and BSF cultures were seeded at $2x10^6$ cells/mL or $1x10^5$, respectively, in biological replicates, counted every other day and diluted back to the above seeding densities every other day for PCF and daily for BSF.

Generation of BSF cells that exclusively express v5-tagged KH2F1 and that lack kDNA

Endogenous KH2F1 alleles were knocked out from BSF strain Lister 427 Single Marker (SM) cells using floxed blasticidin (BSD)- and puromycin (PAC)-Herpes Simplex virus thymidine kinase (HSVTK) drug cassettes as previously described (47,48). Correct insertion of knockout cassettes was assessed by PCR (see **Supplementary Table ST2** for a list of DNA oligonucleotide primers and plasmids). Drug cassettes were excised following each allele knockout via transient expression of Cre recombinase from pLEW100Cre_del_tetO (Addgene plasmid 24019; a gift from George Cross, The Rockefeller University, United States) and selection with ganciclovir (Invivogen, United States) as previously described (11,47). Prior to second allele knockout, the cells were also transfected with a *Not*I-linearized pEnT6+ATPaseGammaWT+3UTR construct that allows the replacement of an endogenous

WT ATPase gamma subunit allele with a mutant allele containing the L262P mutation and selected with BSD (49). Expression of this mutant gamma ATPase (MGA) allele was previously shown to compensate for the loss of the mitochondrial genome and gene expression in BF *T. brucei*, including of RNA editing and editing proteins (11,49-51). Finally, the cells were transfected with *NotI*-linearized pHD1344tub(PAC)-KH2F1-Cterm3V5 plasmid that allows for which allows for constitutive expression of C-terminal 3xV5 tagged KH2F1 from the β-tubulin locus (52). kDNA was removed from the KH2F1 null, KH2F1-v5 BSF cells that contained the MGA allele by treatment with 20 nM acriflavine, maintaining cell density between 1×10^5 cells/mL and 2×10^6 cells/mL (49).

***In vitro*-induced differentiation of monomorphic BSF cells into PCF cells**

In vitro-induced differentiation of monomorphic Lister 427 SM BSF into replicating PCF was performed using reported conditions with some modifications (53). Cells at a 5×10^5 cells/mL density were first treated with 10 µM 8-pCPT-2′-O-Me-cAMP (aka cAMP; Cayman Chemical) in HMI-9 medium at 37°C, 5% CO2. These cells were transferred into SDM-79 medium with 6 mM citrate/cis-aconitate to induce differentiation to replicating PCF at 26°C. Basic hallmark changes in development were verified independently in two labs using common primers for BSF and PCF-specific transcripts and pre-edited and edited mitochondrial transcripts, e.g., major surface antigens and fully edited COX2 (**Supplementary Table ST2**). Initial assays used conventional qPCR. Subsequent assays using Biomark HD Fluidigm examined a broader range of transcripts. Briefly, total RNA was harvested from ~2×10^8 cells using TRIzol pre- and two days post-addition (or not) of 10 µM cAMP, and two, four, or six days after transfer into SDM-79 medium supplemented with 3 mM citrate/cis-aconitate. Samples labeled "d0, d2 or d8 + cAMP" in the main text figures indicate cells prior (control) or two- or eight-days post cAMP treatment, respectively. Isolated RNA was treated with 10 U of Turbo DNase (Life Technologies) according to the manufacturer's instructions and purified through acid phenol:chloroform extraction. A 1 µg aliquot of total RNA was used to synthesize cDNA using the iScript select cDNA synthesis kit with random hexamers. cDNAs were then analyzed directly in qPCR reactions. For conventional qPCR, the cDNA was amplified using the iTAQ SyBr green supermix (Bio-Rad) and primers described in **Supplementary Table ST2** on a 96-well plate using the following thermocycling conditions: 1 cycle at 95°C for 30 s and 36 cycles of 95°C for 5 s, 55°C for 15 s and 60°C for 30 s. Reactions were conducted on a C1000

touch thermocycler plus CFX96 Real-Time System (Bio-Rad). Data were analyzed in the Bio-Rad CFX Manager real-time PCR analysis software (v3.1), using the linear (derivative) baseline correction method and the auto (global) threshold cycle (CT) method. All amplicons were cloned and verified by Sanger Sequencing, and single products were confirmed on a 2% agarose gel. For Biomark Fluidigm assays, the random hexamer generated cDNA was pre-amplified in multiplex specific-target-amplification (STA) reactions using TaqMan PreAmp master mix (Life Technologies) and with the following thermocycling conditions: 1 cycle at 95°C for 10 min and 14 cycles of 95°C for 15 s and 60°C for 4 min. Pre-amplified cDNA was treated with exonuclease I (New England Biolabs) and diluted 5- (for BSF) or 10-fold (for PCF). High-throughput real-time PCR was then conducted on the BioMark HD system with Fluidigm 48-by-48 dynamic array integrated fluidic circuits (IFCs), using SsoFast EvaGreen supermix with Low ROX (Bio-Rad) and primers described in **Supplementary Table ST2**. Primers were as reported (54) or designed in this study. Processing of the IFCs and operation of the instruments were performed according to the manufacturer's procedures. PCR was performed using the thermal protocol GE Fast 96 × 96 PCR + Melt (v2.pcl). Data were analyzed in the Fluidigm real-time PCR analysis software, using the linear (derivative) baseline correction method and the auto (global) threshold cycle (CT) method. For both conventional qPCR and Biomark Fluidigm assays, the CT values determined were exported to Excel software for further processing. Calculations of fold changes in RNA levels in samples following *in vitro* differentiation were done via the 2 [−ΔΔC(T)] method (55) using mtRNP or TERT as an internal reference. We found similar levels of mtRNP mRNA between stages in this study. TERT mRNA was reported to be relatively constant between stages *(56)*. Triplicate replicates of each cDNA sample were assayed for each target and internal reference per experiment, and C(T) data was averaged before performing the 2 [−ΔΔC(T)] calculation (55).

Western blot and immunofluorescence analyses of *T. brucei* editing proteins

SDS-PAGE followed by Western blot analyses of BSF cell lysates to detect KREH2 and KH2F2 (REH2C complex subunits), and RESC13 (RESC complex subunit) were performed as previously described for PCF cell lysates (13,32,38). RESC2 was detected using an anti-RESC2 mouse monoclonal antibody (1:25 dilution) (57), and KH2F1-v5 was detected with rabbit anti-v5 antibody (1:5,000 dilution; Thermo Fisher), followed in each case by anti-mouse anti-IgG secondary antibody (1:5,000 dilution; Bio-Rad). Immunofluorescent microscopy of *T.*

brucei BSF cells was carried out as in our recent study with some modifications (38). Briefly, 10 mL of BSF strain Lister SM 427 cells at ≤ $1x10^6$ cells/mL were pelleted at 2,500 *g*, 4°C for 10 mins, then washed with 1 mL IF wash buffer (1 mM $CaCl_2$, 1mM $MgCl_2$, 1xPBS pH 7.4) and pelleted again. Cells were resuspended in fixing solution (4% paraformaldehyde in 1xPBS pH 7.4) at $1x10^7$ cells/mL final concentration and incubated at room temperature for 5 mins. Fixed cells were applied to a poly-L-lysine coated 4-well chambered coverglass (Thermo Fisher) The slide had been previously treated with 1 M KOH for 1 hour at room temperature, rinsed three times with distilled water, and allowed to air dry. Poly-L-lysine coating was done at 37°C and slides allowed to dry by laminar flow. A 150 mL aliquot of fixed cells (~ $1.5x10^6$ cells total) were pipetted into the slide chambers and allowed to adhere for 10 mins at room temperature. The supernatant was removed by pipetting along the side of the chamber, and chambers were washed by gently pipetting 250 mL of IF wash buffer and incubating for 3 mins. The buffer was then removed and replaced three more times. Cells were permeabilized with IF permeabilization buffer (0.1% Triton-X in 1xPBS pH 7.4) for 10 mins at room temperature before washing with three changes of IF wash buffer. Cells were then blocked with 250 mL blocking buffer (4% BSA in 1xPBS pH 7.4) for 30 mins at room temperature. The blocking buffer was removed by pipetting, and 250 mL primary antibody, diluted in blocking buffer, was individually added into the slide chambers and incubated at room temperature for 45 mins. Dilutions for antibodies were 1:5,000 KREH2, 1:5,000 RESC1/2, and 1:1 KREL1 (i.e., 125 mL blocking buffer and 125 mL of antibody). Slides were then washed with 3 changes of IF wash buffer and stained with 10 mg/mL DAPI for 10 mins then washed with 3 changes of IF wash buffer. Cells were treated with 1:750 diluted Anti-Rabbit IgG Atto 488 (Sigma-Aldrich) and 1:375 AlexaFluor 647 Donkey anti-mouse IgG (Invitrogen) secondary antibody for 1 hour at room temperature. Slides were washed with three changes of IF wash buffer and dried with an air gun. Slides were imaged on a Nikon Eclipse Ti inverted epifluorescence microscope using a 100×objective (Plan Apochromat, NA 1.45, oil immersion), and images were analyzed with NIS-Elements Advanced Research software packages.

Preparation of RNA for library construction, cDNA synthesis, Illumina sample preparation, and sequencing

Total mtRNA or RNA in complexes immunoprecipitated by anti-RESC6 or anti-RESC1 antibodies was isolated from four biological replicates of cells following BSF and PCF ± RNAi induction as previously described (32,38). Gene-specific cDNA synthesis was carried out using oligo 2278 (see **Supplementary Table ST2**), which hybridizes to the 3′ terminus of ND7. cDNA synthesis was carried out with 2 µg of mtRNA using the iScript Select cDNA Synthesis Kit (BioRad). We checked for the specificity of targeted cDNA synthesis and subsequent amplification as follows: BSF and PCF gene-specific cDNAs were amplified using KAPA Hifi HotStart ReadyMix (Roche) using primers containing universal Illumina adapters (5′ 2141/2142, 3′ 1682/2608) (**Supplementary Table ST2**). PCR products were then amplified with primers 2542 and 2543 to generate amplicons with terminal 5′ HindIII and 3′ BamHI sites. These amplicons and pLEW100v5 plasmid were digested with BamHI-HF (NEB) and HindIII-HF (NEB) and gel purified in 1% agarose gels using the NucleoSpin Gel and PCR Clean-up Kit. Digested amplicon and pLEW100v5 plasmid were ligated using T4 DNA ligase (NEB) at a 5:1 molar ratio (insert:vector) to produce plasmid p589 (ND7 3′ domain). The ligated plasmid was transformed into Stellar competent cells (TakaraBio), and 10 individual bacterial colonies were picked and analyzed by Sanger sequencing using primers 2542/2543. Illumina libraries were prepared as previously described with modifications (13). ND7 libraries were amplified from 10 ng BSF or PCF gene-specific cDNA for 24 cycles with primers containing Illumina adapters (ND7 3′ domain 1682/2608). Amplicons containing adapters were purified using Ampure XP PCR purification beads (Beckman-Coulter) and visualized on a 4200 Agilent Tapestation. 10 ng of each amplicon was indexed using the Nextera XT Index Kit V2 SetA (Illumina) with 12-cycles of amplification. The resultant Illumina libraries were then purified again with Ampure XP beads, and their concentration was measured using a QuBit 4 Fluorometer and high-sensitivity dsDNA QuBit Reagent (Thermo Fisher). Libraries were diluted to 4nM and pooled. The library pool was denatured and diluted to 8 pM per the manufacturer's instructions with 40% PhiX spike-in. Samples were run with an Illumina MiSeq Reagent Nano Kit v2 2x250, which produced 1.56×10^6 paired reads after QC filtering and removal of PhiX spike-in.

Processing RNA-seq data in ND7 3′ editing and identifying non-cognate gRNA isoforms that complement the extended 3′ element

Amplicon RNA-seq of ND7 3′ domain editing was processed as reported (13). Subsequently, sample alignment output data were further processed in the R environment (http://www.r-project.org) for summarizing and figure-generation purposes. Searches for gRNAs encoded in the *T. brucei* BSF strain EATRO 1125 minicircle genome (8) that match alternatively edited mRNA sequences were performed using Python scripts (package 3.7) as previously described (9). Alignments of predicted (annotated in minicircles) and sequenced gRNA in EATRO 1125 total mtRNA (8,9) are available online at http://hank.bio.ed.ac.uk. Alignments of sequenced gRNA in PCF strains EATRO 164 total mtRNA (10) and Lister 427 in total mtRNA and RESC6-immunoprecipitations (31) are available online at http://bioserv.mps.ohio-state.edu/RNAseq/T-brucei/MRBs/. We confirmed the sequence of pre-edited ND7 used in this study by PCR amplification of genomic DNA and Sanger sequencing of three independent amplicons. Analyses of last-edit sites were performed as follows. ND7 3′ RNA-seq data was analysed in the R-environment (v4.2.2; R core team 2022; https://www.r-project.org/) to determine the position of the last edited site in the top 100 most abundant amplicons in the corresponding RNA-seq library. Data wrangling was assisted by the "tidyverse" R package (v2.0.0; https://doi.org/10.21105/joss.01686) (58). RNA-seq amplicons were cross-referenced against the ND7 pre-edited sequence one nucleotide at a time, progressing from the 5′ to the 3′ end. The first (5′ most) detected mismatch was classified as the last edit site in that amplicon. Last-edit site frequency was calculated from the top 100 most abundant amplicons by totalling the number of reads with a specific last-edit site and dividing by the total number of reads from the 100 most abundant amplicons.

Sample generation for *in-vitro* DMS-MaPseq

RNA structure of full-length ND7 was experimentally determined *in vitro* by DMS-MaPseq (59). Synthetic gBlocks (IDT DNA Technologies) were generated containing the entire ND7 pre-edited (PE, gBlock 2404) or the most common isoform of ND7 bearing the 3′ High Frequency Element described in this study (HFE, gBlock 2405; **Supplementary Table ST2)**. gBlocks for ND7 were amplified with primers 2631 and 2632, and amplicons were gel purified (Machery-Nagel) and cloned into plasmid pHD1344Tub(PAC) (47), p557 and p558, respectively. Plasmids were verified by Sanger sequencing, and linearized templates with XhoI were used for T7 *in vitro* transcription as reported for another edited mRNA (38). Briefly, 1 µg of purified DNA template was used in the Hi-Scribe T7 transcription kit (NEB) at 37°C

overnight (~16 hr.). DNA template was digested by adding 5U Turbo DNase and incubated at 37°C for 15 mins. Synthesized RNA was isolated using the Zymo Cleanup kit (Zymo Research). 10 µl of 2 µg of purified RNA was denatured at 95°C for 1 min. 89.5 µL 1x Refolding Buffer (397 mM sodium cacodylate buffer (Electron Microscopy Sciences) and 6 mM $MgCl_2$), was added to denatured RNA on ice, and the RNA was allowed to refold at 26°C for 20 mins. 0.5% (v/v) Dimethyl Sulfate (Thermo Fisher) was added, and the reaction was incubated at 37°C for 4 mins shaking at 800 RPM. The reaction was quenched by adding 60 µL 2-Mercaptoethanol, and RNA was cleaned using the Zymo RNA cleanup kit.

DMS-MaPseq libraries were generated using IDT's xGenTM Broad-Range RNA LibraryPrep Kit as reported with slight modifications. Briefly, 500 ng of purified DMS-modified RNA, ND7 PE or ND7 3' HFE, was used as input (3 technical replicates for each condition). 500ng of purified DMS-modified RNA was used as input. Briefly, RNA was fragmented for 2 minutes using the manufacturer's instructions without adding reagent F2 (dNTPs). After 2 minutes, the fragmentation mix was placed on ice immediately. The mixture was then incubated with the reverse transcription mix (1 µl TGIRT (Ingex), 1 µl water, 1 µl enzyme R1 (RNAse inhibitor), and 1µl DTT) at room temperature for 30 minutes. Then, dNTPs (F2) were added, and the fragmented RNA mixture was reverse transcribed under the conditions: 20°C for 10 mins, 42°C for 10 mins, and 55°C for 60 mins, followed by denaturation by adding 1µl 4M NaOH at 95°C for 3 mins. To neutralize the mixture, 2 µl of 4M HCl was added, and the volume was brought up to 50µl with nuclease-free water. Then, the reverse transcribed cDNA was cleaned using a 1x volume ratio of SPRI beads (Beckman Coulter) and eluted in 10µl EDTA TE. Samples were then adapted, extended, ligated, and amplified for 8 cycles for ND7 following IDT's instructions. The libraries (~300-400 bp) were gel-purified on an 8% TBE polyacrylamide gel (Thermo Fisher) and precipitated using isopropanol. To sequence the libraries, samples were loaded on an iSeq-100 sequencing flow cell with iSeq-100 High-throughput sequencing kit and the libraries were run on iSeq-100 (paired-end run, 2x151 cycles).

DMS-MaPseq analysis

To determine the DMS signal, FASTQ files were processed and analyzed using the DREEM (Detection of RNA folding Ensembles using Expectation-Maximization clustering) pipeline

(59). Briefly, reads were trimmed using TrimGalore (github.com/FelixKrueger/TrimGalore) to remove Illumina adapters. Trimmed paired reads were then mapped to *T. brucei* EATRO 1125 maxicircle genome (accession: MK584625) (9) using Bowtie2 with the parameters: --local --no-unal --no-discordant --no-mixed -L 12 -X 1000. A bit vector was generated for each pair of aligned reads, and the mutational signatures were analyzed using the DREEM algorithm (59). The ratio of mismatches and deletions to total coverage at each nucleotide position was calculated to quantify the population average DMS reactivity at each position. DMS reactivities were normalized to the median of the top 5% of DMS reactivities to a scale of 0 to 1. These normalized DMS reactivities were used as folding constraints for predicting RNA secondary structures with the program RNAstructure v.6.0.1 (60). RNA secondary structures were visualized using VARNA v.3.93 (61).

Isolation of *in vivo* chimeric molecules of gRNA gCOX3 with mRNA ND7 or mRNA COX3

To isolate gRNA/mRNA bimolecular chimeras *in vivo*, we generated ND7 or COX3 gene-specific cDNA using 2 µg of DNase-treated mitochondrial RNA from wild-type PCF cells. cDNA was generated using the BioRad iScript Select cDNA synthesis kit per the manufacturer's directions with oligo 2278 (ND7) or 2732 (COX3) in a 20 µL reaction. 2 µL of resulting cDNA was used in a 50 µL Phusion HF polymerase PCR reaction using the forward oligo 2882 (COX3 gRNA) and 2278 (ND7 mRNA) or 2732 (COX3 mRNA) to enrich for chimeras. The resulting PCR product was verified on 2% agarose gel, and the corresponding bands were gel-eluted using the Nucleospin PCR cleanup kit. 10 ng of purified PCR product was used as template in subsequent PCR amplification using nested primers, which added 5' HindIII and 3' XhoI sites for cloning into pHD plasmids (2880/2881 COX3-ND7 and 2880/2883 COX3-COX3). The undigested PCR product was cloned into HindIII/XhoI digested pHD1344Tub(PAC) plasmid using the In-Fusion cloning kit at a 2:1 molar ratio. This reaction generated plasmids p658 (COX3-ND7) and p659 (COX3-COX3). 2 µL of In-Fusion product was then transformed into Stellar competent cells per the manufacturer's instructions, plated on Ampicillin-selective LB agar plates, and allowed to grow overnight at 37°C. Ten individual colonies were picked from each plate (COX3-ND7 or COX3-COX3), and plasmid was isolated from cultures using a QIAgen mini plasmid isolation kit. 25 ng of isolated plasmid was used as a template in a 50 µL Phusion HF polymerase PCR reaction with primers 2056/2242 to

amplify a portion of the vector containing our cloned fragment. These PCR amplicons were then purified by 1% agarose gel electrophoresis, and 150 ng of PCR product was analyzed by Sanger sequencing in technical replicates by MCLAB using oligo 2611 and 2622. Alignments of chimera sequences were done using the MUSCLE alignment tool.

Calculations and statistical analysis

Total editing and NC/C (non-canonical/canonical) ratio values (previously termed Inc/Cor ratio) of the percentage of reads, both site-by-site and cumulative, were calculated as previously reported (13,38). This updated nomenclature reflects a key observation here and in a recent study that specialized non-canonical editing is developmentally controlled by KREH2 (38). Graphs compare replicate sets for two conditions of either total mtRNA or RESC6- or RESC1-RNA immunoprecipitations (RIPs) (e.g., PCF vs. BSF or -Tet vs. +Tet), where one replicate set includes at least three biological replicates, and another set includes at least two biological replicates. These replicate sets enabled statistical calculation of p-values, average, and standard deviation. A description of samples and p-values for all sets compared are included (**Supplementary Table ST3).** Significant changes in editing fidelity (NC/C value) in each site were determined as before (13,38). To generate *P* values for the effects of KH2F1-RNAi on ND7 3′ editing, we combined independent biological replicates for 3- and 4-days post-induction of RNAi since these two type points of treatment had comparable outcomes. This allowed us to increase the total number of replicates for the +Tet condition. We used one-way ANOVA to test the null hypothesis that there is no significant difference between groups, with this null hypothesis rejected at $P < 0.05$. The mean + SD of independent biological replicates was reported.

RESULTS

KREH2 co-localizes with kDNA and associates with KH2F1 and KH2F2 in REH2C complexes of BSF cells

Most editing studies have been performed in PCF cells because they grow at higher densities and generate more material than BSF cells. Our prior studies showed that KREH2 is essential for growth and editing in both PCF and BSF *T. brucei* (38,41,43). We also showed that KH2F1

(REH2C) co-localizes with kDNA (38) and confirmed KREL1 (RECC) and RESC1/2 (RESC) co-localization with kDNA (43,62). We now show for the first time that KREH2 co-localizes with kDNA in BSF cells (**Fig. 1A; Supplementary Fig. S1A**). Furthermore, KREH2 interacts in an RNA-independent fashion with KH2F1 and KH2F2 in a BSF dyskinetoplastic (DK) mutant strain that lacks kDNA and is thereby devoid of mtRNA (49) (**Fig. 1B; Supplementary Fig. S1B**). Thus, RNA-free versions of the REH2C complex exist in BSF as originally proposed in PCF (32).

KREH2 knockdown differentially affects total editing in ND7 3′ domain in PCF and BSF

Prior qRT-PCR assays confirmed that KREH2 promotes accumulation of mature fully edited mRNAs in PCF and BSF *T. brucei* (38,41,43). However, these assays do not inform the precise details of 3′-5' editing progression. Here, we applied targeted base-resolution RNA-Seq to examine how KREH2 knockdown in PCF and BSF may affect the editing progression of the 3′ terminus in ND7 3′ domain (aka ND7 3′), including ORF (90 sites) and UTR (7 sites) sequences in the amplified fragment (**Supplementary Fig. S2A**). We began by examining ND7 3′ in total mtRNA. To illustrate the raw data initially collected and tallied by our bioinformatics pipeline, we included a stack plot of representative biological replicates in PCF and BSF cells (**Figs. 2A-B**). These plots provide a snapshot of all editing events scored at each site in the ND7 3′ amplicon.

The total editing action on an mRNA comprises two separate components, canonical and non-canonical. Changes between these two types of edits were not necessarily reciprocal in previously examined A6 and RPS12 mRNA substrates (13,38). We focused on the 3′ region of ND7 3′ because our plots exhibited extensive editing action and dramatic differences along the first few blocks between BSF and PCF (**Figs. 2A-B**). The first gRNA, initiator gRNA-1, covers most of the 3' UTR and also two sites that create the stop codon. Editing is highly efficient along these first few sites in both PCF and BSF. However, the region spanned by the 3′ terminus of gRNA-1 (e.g., sites 31 and 33) and subsequent guides exhibited evident differential changes. In BSF, editing action, mostly canonical, remained high along the first five gRNAs but dropped dramatically near the 5′ end of the examined fragment. In contrast, in PCF, editing action gradually decreased, starting at the first sites potentially targeted by gRNA-2. PCF cells exhibited relatively low canonical editing along most guides examined, but

we detected substantial partial non-canonical editing (i.e., yellow and black bars), particularly opposing the 3′ end of gRNA-1 and across gRNA-2. We hypothesized that early editing across the first two guides might offer early differential checkpoints in ND7 maturation during development, which we examined in more detail.

In the ND7 3′ fragment amplicon, the forward primer matched a pre-edited sequence, which likely selects for the low editing action at the 5′ end of the sequenced examined. Because the mRNA 3′ terminus is short and AU-rich, a suitable reverse primer was used to tally U-indels beginning at the second position for canonical editing (site 21) in ND7 3′ (**Supplementary Fig. S2A**). The initiator gRNA-1 in our alignments represents gRNA-1 gND7(1269-1320) (22), aka gND7 *B1 (31)* in PCF strains EATRO164 and Lister 427, respectively. gRNA-2 in our alignments is one of two reported potential guides, gND7(1240-1268) (22), aka gND7 *B2.alt (31)* in PCF strains EATRO164 and Lister 427, respectively. We reported that gND7 *B2.alt* best matches fully-edited ND7 3′ examined by Sanger-sequencing in PCF strain Lister 427 (31,42). These first two guides also generate the best match with the ND7 3′ canonical pattern in our BSF samples. In BSF strain EATRO1125, the annotated guides in minicircle DNA libraries predict the same ORF in ND7 3′ except that it ends in VDR (9) (**Supplementary Fig. S2B**). This is one of two alternative ORFs proposed in EATRO164 ending in VDR or EYR (22). The alternative sequence forming EYR was present in up to ~70% and ~14% of all reads in BSF and PCF in Lister 427 in this study, in line with our early sequencing studies (31,42). The alternative sequence forming VDR was present in 0.2-1% of all reads in our samples, and its frequency was not evidently affected by cell stage or KREH2 expression (see the statistics in **Supplementary Table ST3**). Thus, canonically edited ND7 3′ with a predicted ORF ending in EYR may be the "functional version" in strain Lister 427 under control during development.

All our subsequent analyses directly compared independent biological replicates of each sample plus or minus KREH2 knockdown. We focused on a region that spans the first two guides in PCF, particularly sites 33 and 37-to-43, which includes extensive non-canonical editing. However, analyses of entire amplicons were also included (**Supplementary Figs. S3-S7**). We plotted total editing at each site as well as cumulative editing, either along the amplicon or up to the most 5′ site examined. We compared total mtRNA in PCF and BSF and RNA-immunoprecipitations (RIPs) of RESC6 and RESC1 in PCF. Among many potential RESC isoforms (30,31,63), RESC-A contains RESC6 and RESC1, while RESC-B contains RESC6 but not

RESC-1 (36). This is the first time isolated native RESC1 complexes have been examined by RNA-Seq, as was previously done for native RESC6 complexes (13,38). Total mtRNA samples without KREH2 knockdown confirmed significant differences between PCF and BSF in analyses of total editing action at most ND7 3′ sites examined (**Fig. 2C; Supplementary Fig. S2C**). As expected, cumulative total editing at the most 5′ site exhibited higher editing action in BSF than PCF (**Fig. 2D**). Also, total ND7 3′ editing action was significantly enriched in both RESC6 and RESC1 RIPs vs. total mtRNA (**Fig. 2E**). This was expected since our earlier RT-qPCR studies had shown enrichment of fully-edited and pre-edited ND7 3′ in RESC6-RIPs vs. total mtRNA (31).

KREH2 knockdown by ~80% (32,38) significantly decreased total editing in mtRNA and RESC6-RIP samples. This effect was observed in most sites examined (**Figs. 3A-B; Supplementary Figs. S3A-C**) and in graphs of cumulative total editing up to the most 5′ site (**Figs. 3C-D**). RNAi-knockdown of KH2F1, another subunit of REH2C, confirmed the above effect on total editing in PCF mtRNA and RESC6-RIPs (**Figs. 3C-D**; **Supplementary Fig. S3D**). We previously reported that KH2F1 knockdown in PCF destabilizes KREH2, effectively creating a dual knockdown (32). Thus, specific KREH2 knockdown, which does not affect the KH2F1 integrity (32), suffices to decrease ND7 3′ total editing in PCF. Cumulative total edits were similar in RESC1 and RESC6-RIPs, but total editing inhibition by KREH2 knockdown was more robust in RESC1-bound complexes (**Fig. 3D)**. These results are in line with prior qRT-PCR assays showing that REH2C loss-of-function reduced the steady-state accumulation of mature fully edited ND7 3′ in PCF and BSF (13,32,38,42).

KREH2 knockdown differentially affects ND7 3′ editing fidelity and pausing in PCF and BSF

Prior studies established that KREH2 knockdown decreased editing fidelity in mRNAs RPS12 and A6 in substrate and site-specific fashion (13,38). We score editing fidelity using the normalized NC/C ratio: the percentage of non-canonical reads over canonical reads (**Supplementary Table ST1**, glossary of terms). Here, we asked whether KREH2 affects editing fidelity in ND7 3′, particularly across the first two guides, where dramatic differences in editing action were apparent between PCF and BSF. Large NC/C values indicate low

editing fidelity in site-by-site and cumulative plots. As mentioned above, the canonical sequence examined here likely represents a functional ND7 3′.

Remarkably, KREH2 or KH2F1 knockdowns in PCF increased editing fidelity (i.e., significantly reduced the NC/C ratio) in ND7 3′. This effect was observed at most sites examined, including sites 33 and 37-43, and in cumulative NC/C charts in mtRNA and RESC6-RIPs (**Figs. 4A-D; Supplementary Figs. S4A-D; Figs. S5A-C)**. Generally, increased editing fidelity in the knockdowns involved concurrent downregulation and upregulation of non-canonical and canonical edits, respectively (**Figs. 4E-F; Supplementary Figs. S6-S7).** Furthermore, this increase in editing fidelity involved more robust changes in non-canonical than canonical editing and was more evident in RESC6-RIPs than mtRNA. These results suggest that REH2C actively modulates editosome function.

In other comparisons, cumulative NC/C in ND7 3′ was approximately five-fold higher in RESC6-RIPs and mtRNA than RESC1-RIPs (**Fig. 4D)**. However, KREH2 knockdown also increased editing fidelity in RESC1-RIPs. This suggests that RESC6-containing RESCs are a major platform for KREH2-dependent low-fidelity ND7 3′ editing in PCF. Also, although ND7 3′ editing fidelity was generally higher in BSF than PCF, KREH2 knockdown also decreased the NC/C ratio at several sites in BSF. However, knockdown of KREH2 in BSF did not upregulate canonical edits as in PCF. In BSF, both non-canonical and canonical editing decreased in the KREH2 knockdown, but in different proportions, i.e., the former declined significantly, while the latter decreased moderately. Combined, these changes reduced the cumulative NC/C ratio in BSF (**Figs. 4E-F)**. Importantly, KREH2 knockdown in BSF did not significantly affect the NC/C ratio at sites 33 and 37-43 **(Supplementary Figs. S4B, S4E)**.

Sites where canonical editing naturally decreases are known as intrinsic pause sites (IPSs) (12). Peaks in NC/C indicate pausing in canonical editing from one site to the next. That is an interruption in canonical progression due to persisting alternative edits upstream (13). In PCF, site 33 had the largest examined NC/C value in ND7 3′, and IPS32 was the largest pause site identified, i.e., NC/C at site 33 was 107-fold and 160-fold higher than at site 32, in RESC6-RIPs and mtRNA, respectively (**Figs. 4A-B; Supplementary Figs. S4)**. Thus, canonical editing may continue past IPS32 in ~1 out of 107 or 160 molecules in RESC-RIPs and mtRNA, respectively. As mentioned above, the KREH2 knockdown reduced the NC/C at most sites

examined, including site 33, thereby alleviating IPS32 to facilitate editing progression. In typical IPSs, canonical editing resumes at the following upstream site. However, our bioinformatic analyses revealed tracks of sites, each site exhibiting high NC/C, that is, an "intrinsic pause track," including sites 37-43. KREH2 knockdown alleviated this and other intrinsic pause tracks in ND7 3′ in mtRNA and RESC6-RIP samples. Among others, particularly high NC/C values in ND7 3′ identified IPS35 and IPS36 next to the 37-43 site track in block 2 (by gRNA-2), and IPS48 and IPS54, next to sites 49 and 55, respectively (**Figs. 4A-B; Supplementary Fig. S4)**. Major IPS stimulated by KREH2 were termed KREH2-promoted pause sites (PPSs). As indicated above, IPS32 may be the largest KREH2-PPS in ND7 3′.

In brief, PCF-specific repression of ND7 3′ mRNA editing involves decreased editing fidelity and increased pausing in this sequence. This repression involving high NC/C ratios, particularly in early editing, requires KREH2 and seems more effective in native RESC6-bound RESC complexes. KREH2 is the first factor identified that naturally hinders editing fidelity and progression in a specific substrate in PCF.

Opposite modulation of non-canonical and canonical editing by KREH2 reveals a major early checkpoint in PCF

To further examine the source of ND7 3′ low editing fidelity (high NC/C ratios) in PCF, we surveyed all editing events, particularly at sites 33 and 37-43, near the ORF 3′ end (**Figs. 5A-B**). Analyses of all non-canonical reads revealed dominant events at these sites. At site 33, a +2U insertion event in RESC6-RIPs accounted for 94% of the NC value (~60% of all reads, i.e., canonical, non-canonical, and pre-edited). Dominant non-canonical reads at sites 37-42 accounted for ~80 to 90% of the NC value (>40% of all reads) and slightly less at site 43 (**Fig. 5C**). These non-canonical reads were also dominant in RESC1-RIPs but less frequent (<10% of the NC value; < 5% relative to all reads) (**Fig. 5C**). Interestingly, the +2U insertion at site 33 might be directed by two encoded 3′-terminal adenines in gRNA-1 (**Fig. 5B**). These adenines are not part of the classified canonical guiding domain in gRNA-1 but are conserved in the sequenced gRNA-1 in strains Lister 427 (used in this study), EATRO 164, and EATRO 1125 (**Supplementary Fig. S2B**) (9,22,31). We previously reported a dominant non-canonical +2U insertion in mRNA RPS12 potentially directed by two conserved 3′-terminal

adenines of gRNA-1. A +2U insertion by conserved 3′-terminal adenines in gRNA-1 in distinct substrates suggests that this event is biologically important.

The proportion of dominant non-canonical reads at the cluster of sites that include 33 and 37-43 decreased in RESC6-RIPs (*P*= 0.006) and mtRNA (*P*= 0.02) upon KREH2 knockdown. While the +2U event at site 33 was by far the most abundant non-canonical event in ND7 3′, we hypothesized that dominant non-canonical reads at sites 33 and 37-43 may co-exist in a sequence element in ND7 3′. To test this idea, we searched for a consensus element in two versions, including sites 33-43 (long) and 37-43 (short) (**Fig. 5D**). Since sites 34-36 have variable U content, they were allowed any U number. These searches revealed a 3′ High-Frequency Element (3′ HFE) in ~30 % of ND7 3′ amplicons in all PCF samples without KREH2 knockdown (**Fig. 5E,** left panels -Tet). The long and short elements exhibited very similar frequencies. Over 99% of the hits with the short form contained the long form. This suggested that the dominant non-canonical events at sites 33 and 37-43 were concerted.

We then asked whether KREH2 affects the frequency of the 3′ HFE in ND7 3′. Knockdown of KREH2 in PCF significantly decreased the 3′ HFE reads in mtRNA and RESC6-RIPs (**Fig. 5E**, left panels). KH2F1 knockdown in PCF confirmed a decrease in 3′ HFE reads in mtRNA. In RESC6-RIPs, the average 3′ HFE reads appeared to decrease upon KH2F1 knockdown, but this change was not significant, presumably due to variation between our KH2F1-RNAi sample replicates. Because the 3′ HFE is installed in the same region for canonical editing by gRNA-2, we asked whether 3′ HFE formation affects block 2 maturation. Remarkably, knockdown of KREH2 or KH2F1 significantly increased block 2 maturation in PCF mtRNA and RESC6-RIPs (**Fig. 5E**, right panels). To our knowledge, KREH2 is the first identified protein that specifically represses the maturation of an entire editing block in any substrate. Other blocks were examined below. KH2F1 may participate at least by affecting KREH2 stability in PCF (32).

To better understand how KREH2 hinders block 2 maturation in PCF, we compared the percentage of 3′ HFE, canonical, pre-edited, and remaining non-canonical "partial" editing reads in the block 2 region (**Fig. 5F**). The pool of partial reads may include related 3′ HFE variants that did not exactly match the consensus element in our searches, plus unrelated sequences of unclear origin. We calculated the percentage of partial reads by subtracting the total reads minus the sum of other read types in the block 2 region: pre-edited, canonical,

and consensus 3′ HFE. In RESC6-RIPs in PCF, pre-edited reads were the lowest fraction (13%), consistent with these samples representing active RESC variants. Among read types exhibiting any editing action, partial edits were the highest (36%), followed by 3′ HFE (28%) and canonical reads (23%). Upon KREH2 knockdown, canonical reads became the highest (33%) and 3′ HFE the lowest (23%). Partial reads decreased (30%), while pre-edited reads were minimally affected, remaining the lowest (14%) in the knockdown. Thus, during early ND7 3′ editing in native RESC6 complexes, KREH2 mostly affected the ratio of edited product types (non-canonical vs. canonical sequence) rather than the overall amount of substrate consumption. In total mtRNA in PCF, KREH2 knockdown also decreased 3′ HFE and partial reads while increasing canonical and pre-edited reads. In RESC1-RIPs in PCF, 3′ HFE reads were rare (3%) and minimally affected upon KREH2 knockdown. Notably, in BSF, the 3′ HFE mtRNA was negligible (0.01%), consistent with the previously observed efficient ND7 3′ maturation in this lifecycle stage. Thus, the 3′ HFE was, on average, ~3,000-fold enriched in PCF vs. BSF in ND7 3′ amplicons from mtRNA. We also asked whether KREH2 knockdown promotes the maturation of other editing blocks in ND7 3′. Importantly, in RESC6-IPs, KREH2 knockdown significantly increased the maturation of blocks 3 and 4 but not block 1 (**Fig 6A)**. We found similar results in mtRNA samples, except that block 1 maturation decreased upon KREH2 knockdown. Our results in the examined blocks 2 to 4 align with the global observed changes in the NC/C ratio and the individual NC and C values at most sites in ND7 3′ (**Supplementary Figs. S6-S7)**. Thus, KREH2 may modulate the selection of gRNAs, cognate vs. non-cognate, along most of the ND7 3′ sequence examined specifically in PCF.

These studies so far have compared PCF vs. BSF (monomorphic) cell lines of strain Lister 427. However, these lines have been cultured separately for over a decade. A central question is whether the 3′ HFE accumulates in PCF during development. We addressed this question using *in vitro*-induced differentiation of cultured BSF monomorphic cells into replicating PCF using reported conditions (53). Conventional RT-qPCR showed the expected changes of major surface antigens during differentiation, i.e., increased procyclin, and massive loss of variant surface glycoprotein mRNAs (53). We also confirmed a robust PCF-specific accumulation of fully-edited COX2 (i.e., activation of oxidative phosphorylation) (**Fig. 6B)** (64). These and additional changes in a broader set of transcripts during differentiation were confirmed by Biomark Fluidigm assays, including a robust increase of fully-edited COX3, and

CYb transcripts, and a moderate increase of nuclearly encoded COX5, COX6, and COX10 (**Supplementary Fig. S8A**). To assess 3′ HFE accumulation in these cells, we established an RT-qPCR assay of 3′ HFE to quantitate this element relative to canonically edited and pre-edited ND7 3′ (**Figs. 6C-D**; **Supplementary Fig. S8B**). We conducted independent differentiation experiments in two labs and scored transcripts by qPCR normalized to mt-RNA polymerase (mt-RNP) mRNA or telomerase RNA transcript (TERT), which we found (mt-RNAP) or was reported (TERT) to be relatively stable between trypanosome stages (56).

Notably, these two independent RT-qPCR measurements confirmed relatively higher or lower 3′ HFE and ND7 3′ canonical editing in parental PCF vs. BSF cells, respectively. *In vitro* differentiated PCF (±cAMP + CCA conditions) exhibited partially increased 3′ HFE and decreased ND7 3′ canonical editing in both experiments. Further normalization of these transcripts to pre-edited substrate, which was elevated in PCF vs. BSF, also confirmed the expected changes in 3′ HFE and ND7 3′ canonical editing upon differentiation.

Thus, *in vitro* differentiation recapitulated the opposite changes in 3′ HFE and canonical editing observed in PCF vs. BSF parental cells and indicated that these changes represent a bonafide phenomenon during development. We note that the detected changes in these transcripts by RT-qPCR were generally more modest than those captured in Ilumina samples. This discrepancy between the two quantitative approaches may partly reflect differences in primer binding efficiency for the amplicons examined in our RT-qPCR conditions.

Overall, in PCF, KREH2-mediated opposite modulation of canonical and non-canonical editing may repress the maturation of entire blocks in ND7 3′. KREH2 promotes 3′ HFE formation in ~30% of the ND7 transcriptome examined in Illumina samples. This event, combined with the inhibition of block 2 canonical editing, reveals a major checkpoint in early editing. However, we also found KREH2-dependent opposite control of canonical and non-canonical editing in most of the ND7 3′ editing examined in PCF. In BSF, the 3′ HFE is much less frequent and not modulated by KREH2. Native RESC6 complexes generally exhibited more efficient KREH2-negative control of ND7 3′ than total mtRNA or native RESC1 complexes. This form of control in ND7 3′, including 3′ HFE formation in early editing, was reproduced upon *in vitro* differentiation.

3′ HFE formation involves the initiator gRNA-1 and a novel terminator gRNA and abolishes upstream canonical editing

To better understand how the abundant ND7 3' HFE is created and may impact editing progression, we initially examined the top 10 most common 3' HFE-containing sequences in RESC6-RIPs ± KREH2-RNAi. Notably, in most amplicons, all editing action had ceased upstream of the 3′ HFE (**Fig. 7**) after a short non-canonical editing junction of variable composition. Some amplicons lacked a junction entirely and were pre-edited immediately upstream of the 3′ HFE, i.e., junction length (JL) = 0. All samples without KREH2 (or KH2F1) knockdown had the same top three amplicon species, including one with JL= 0 (**Fig. 7A; Supplementary Figs. S9-S11**). Tallies of the top 100 amplicons confirmed the most common type with JL= 6 (**Fig. 7A**; the last edit site in each junction was tallied, and its percentage was given above the sequence). KREH2 (or KH2F1) knockdown generally increased the junction length in 3' HFE-containing transcripts, and the species with JL = 0 was lost in the top ten amplicons in most samples (**Figs. 7B-C; Supplementary Figs. S9-S11**).

The 3' HFE would prevent anchoring by upstream cognate guides. Accordingly, bioinformatic searches showed that 3' HFE-containing amplicons in our samples lacked canonical editing of blocks 2-to-4. Notably, the 3' HFE abuts a fully edited 3' terminal sequence in mature ND7 3′ (13,42). Combined, this 3′ terminal canonical sequence plus the 3' HFE make a consensus extended element through site 43 (≥54-nt): 5′-AuuuuuGuAuuuGuuuGuuGuuGuuACCGuuAGAuCGuuAGuuuuAuAuGuGAG-3'. Multi-sequence alignments of the top 100 amplicons bearing the 3' HFE confirmed the presence of the ≥54 nt extended element with minor differences. This observation suggests that this extended element derives from concerted guiding events. We noted that the 3′ HFE length suits the combined average sizes of the guiding and anchor regions in a typical gRNA (i.e., 20-to-40 nt and 6-to-11 nt long, respectively; (8,9). A search in the annotated minicircle genome from *T. brucei* BSF strain EATRO 1125 (8,9) identified a gRNA that matches the 3′ HFE (**Fig. 8A**; transcript #1). Notably, this gRNA anchor would bind, via Watson-Crick base-pairing, to a sequence created by the non-canonical +2U at site 33, while its guiding region would direct the remaining 3′ HFE sequence.

Surprisingly, this gRNA is the previously classified cognate gRNA mO_235(Ia)_ gCOX3(206-250), termed hereon gCOX3, that directs editing in mRNA COX3 (**Fig. 8B**) (8,9). Similar gCOX3 guides are found in other databases in PCF strains EATRO 164 (10) (transcripts #2-to-4) and Lister 427 (transcripts #5 and #6). Equivalent gCOX3 gRNAs were also found in PCF strains TREU 667 and TREU 927 (Donna Koslowsky, personal communication). Most of these guides (transcripts #1-to-5) have a single C:A mismatch with a cytosine dimer in mRNA. This C/A polymorphism and mismatch could potentially cause a non-canonical +1U insertion at sites 34 or 35. However, these sites exhibited relatively low NC/C values (**Fig. 4A**), suggesting that the alternative +1U edit is rare. One isoform (transcript #6) completely matches the 3′ HFE but lacks information for sites 42-43. The conservation of gRNA isoforms in multiple *T. brucei* strains and their complementarity to the extended 3' element suggests that the proposed concerted non-canonical action by the cognate initiator gRNA-1 and non-cognate gRNA gCOX3 is biologically relevant in ND7 3′.

To assess whether gRNA relative frequency may influence the observed levels of 3′ HFE, we examined available Illumina sequenced gRNA-1 and gRNA-2 for ND7 3′ and gCOX3 in several strains (**Fig. 8C**). In PCF strain EATRO 164 mtRNA, the relative abundances of gRNA-2 and gCOX3 isoforms were similar to each other but higher than that of gRNA-1. All three gRNAs appeared to have higher abundances in this strain in BSF vs. PCF. In BSF strain EATRO 1125 mtRNA, these three gRNAs exhibited comparable frequencies. In PCF strain Lister 427 mtRNA and RESC6-RIPs, gRNA-1 was detected in all libraries but appeared relatively enriched in native RESC6 complexes. Interestingly, gCOX3 isoforms were only detected in KREH2-RIPs. In summary, the relative frequency of gRNAs did not explain the observed differences in 3′ HFE abundance between our samples.

We also asked whether the stability of gRNA/mRNA pairing may influence the observed levels of ND7 3′ HFE. Predicted ΔG values of continuous duplexes suggested that the pairs between gCOX3 isoforms and noncognate 3′ HFE-bearing mRNA (**Fig. 8A)** are thermodynamically less stable than pairs between gCOX3 (**Fig. 8B**) or gRNA-2 (**Fig. 5A**) and their cognate mRNAs. However, the efficiency of editing directed by gCOX3 is significantly higher than that directed by gRNA-2 in ND7 3′ in PCF ($P<0.005$) (**Figs. 5E-F;** compare 3′ HFE vs. block 2 edited samples without KREH2 or KH2F1 knockdown). Thus, gRNA/mRNA pairing stability did not explain the observed differences in 3′ HFE abundance between our samples.

These observations, together with the PCF-specific KREH2-dependent upregulation of 3′ HFE, argue against random off-targeting of mO_235(Ia)_gCOX3(206-250) isoforms on ND7 3′ but rather suggest a fixed event in evolution.

To further examine the *in vivo* action of the proposed novel gCOX3, we looked for covalently formed bimolecular chimeras between gCOX3 and ND7 mRNA. Chimeras validate on-target gRNA-mRNA pairs *in vivo* but are not considered true editing intermediates (65-68). We readily isolated gCOX3 chimeras with mRNAs ND7 and COX3 (**Supplementary Fig. S12**). Chimeras with both mRNAs appeared to use the same gCOX3 isoform (#5) in strain Lister 427, based on their sequenced 3′ terminus. Thus, the novel gCOX3 may be bifunctional, that is, playing a dual role, canonical and non-canonical, in COX3 and ND7 3′ editing, respectively. Taken together, these data indicate that ND7 3′ editing has a PCF-specific early control checkpoint where KREH2 normally promotes the use of the inhibitory non-cognate gCOX3 over the cognate gRNA-2. Accordingly, KREH2 knockdown reverses this preferential guide selection. The observed differences in the ND7 3′ HFE level did not correlate with relative gRNA abundance nor with duplex stability of mRNA/gRNA pairs. KREH2 in PCF may normally promote gCOX3 use, introduce a constraint on canonical gRNA-2 function, or both. Either way, the KREH2 knockdown in PCF resulted in a partial relief in overall negative control with a consequent increase in canonical editing at the early checkpoint.

The 3′ HFE may form "repressive" RNA structure in ND7 3′ early editing

Targeting by mO_235(Ia)_ gCOX3(206-250, covers most sites typically modified by cognate gRNA-2 at the 3′ terminus of the ORF in mRNA ND7. However, the 3′ HFE sequence itself may not disrupt downstream binding and subsequent editing by gRNA-2 (**Fig. 7A**). Potential gRNA-2-directed "repair" editing could remove the 3′ HFE and replace it with the canonical block 2 sequence. Because of the 3′ HFE's remarkable abundance, we considered that the 3′ HFE may form RNA secondary structure that somehow hinders gRNA-2 usage. To address this possibility, we applied dimethyl sulfate (DMS) mutational profiling with sequencing (DMS-MaPseq) to examine the structure of mRNA ND7, as in our prior study (38). This probing strategy uses DMS that specifically labels the Watson-Crick face of open and accessible adenine and cytosine bases in the RNA (9,22), and a high-fidelity processive thermostable group II intron reverse transcriptase (TGIRT) enzyme (69). We compared full-

length *in vitro* T7-transcribed ND7 in pre-edited form or the most frequent 3′ HFE-containing molecule in all samples (**Fig. 7; Supplementary Figs. S9-S11**). We used the DMS reactivities as folding constraints to generate structural models of the ND7 3′ terminus (**Figs. 9A-B**). These models showed that the entire 3′ HFE and the binding site for cognate gRNA-2 in the ≥54 nt extended 3′ element form a semi-continuous helical region. This proposed stable fold may reduce access to the binding site for gRNA-2 and thereby "repair" editing by this guide. In contrast, equivalent positions in pre-mRNA appear largely accessible to DMS reactivity. DMS reactivity profiles showed particularly low signals in the predicted fold, including the 3′ HFE (sites 33-43) and binding site for the gRNA-2 anchor, supporting the hypothesis that these features are occluded in ND7 3′, whereas relatively higher DMS reactivity in upstream pre-edited and downstream edited sequence, including the binding site for the initiator gRNA-1, are generally more accessible to DMS. Therefore, our experimentally determined structures support a model whereby KREH2-promoted non-canonical action by the non-cognate terminator gCOX3 and 3′ terminus of the cognate initiator gRNA-1 might form an RNA fold that sequesters the mRNA 3' terminus. The proposed allosteric blockage of downstream cognate guiding would prevent the removal of the 3′ HFE to restore editing progression.

DISCUSSION

ND7 3′ editing mediated maturation is typically more efficient in BSF than in PCF cells, including upon induced *in vitro* differentiation from PCF into mammalian infective forms (14,44,46). However, the key regulatory factors driving stage-specific or preferential RNA editing control have remained a mystery. This study supports a stage-specific editing repression model for ND7 3′ that requires KREH2 (**Figure 10**). In this model, KREH2 concurrently downregulates canonical editing and upregulates gRNA-programmed non-canonical editing in ND7 3′ in PCF cells. Remarkably, KREH2 RNAi-knockdown in PCF relieved repression by reversing the two editing types and confirmed the role of KREH2 in PCF-specific ND7 3′ negative control. KREH2 is the first identified protein to repress canonical editing in any mRNA substrate during development. This negative control by KREH2 occurred at most sites examined in ND7 3′ and included both editing by non-cognate guides

and "unexpected" editing by cognate guides. However, the PCF-specific negative control preferentially occurred at a major early checkpoint identified in this study. At this checkpoint, KREH2 stimulates coordinated non-canonical editing by 3′-terminal bases in gRNA-1 and a novel terminator gRNA, mO_235(Ia)_gCOX3(206-250). Combined, these non-canonical edits install a 3′ HFE in ~30% of the ND7 transcriptome in PCF. This structural element may occlude the mRNA 3′ terminus. The lack of this repression system in BSF facilitates efficient ND7 3′ maturation, i.e., the 3′ HFE is rare (~0.01%), and KREH2 promotes neither 3′ HFE formation nor other non-canonical editing in most sites examined in ND7 3′ in BSF. Critically, *in vitro*-induced differentiation from BSF into PCF confirmed that the opposite change 3′ HFE and canonical editing level in ND7 3′ in parental PCF vs. BSF cell lines is a bonafide phenomenon during development.

Thus, developmental editing control may involve KREH2-dependent selection of gRNAs, that is, housekeeping cognate vs. regulatory non-cognate. Modulated gRNA selection and usage by key editing proteins may help remodel the mRNA ND7 transcriptome during development.

A PCF-specific major checkpoint entails KREH2-dependent modulation of cognate and non-cognate gRNA usage that inhibits early ND7 3′ editing

The major checkpoint ND7 3′ early editing may involve constraints on cognate guiding and or enhanced non-cognate guiding that out-competes the former. A two-step model of KREH2-promoted non-canonical editing by the initiator gRNA-1 and a novel repressor gRNA may install the 3′ HFE. In the first step, two 3′ terminal adenosines in gRNA-1 abutting the classified guiding region direct the most frequent non-canonical edit (+2U at site 33) examined in ND7 3′ (**Supplementary Fig. S6**). In native RESC6-complexes, this +2U event accounted for ~94% of all non-canonical reads (~60% of all reads at site 33). This +2U event led to the most robust KREH2-promoted editing pausing at site 32 (PPS32) in the ND7 3′ sequence examined in PCF. Paradoxically, strong pausing at site 32 requires highly efficient block 1 maturation (**Fig. 6A**). Restoring canonical block 2 editing during development in BSF cells would necessitate "alleviating" PPS32 by reducing the +2U edit by gRNA-1, promote cognate gRNA-2 function (over the non-cognate terminator) to replace the 3′ HFE with canonical sequence, or both.

Like ND7 3′, the initiator gRNA-1 in mRNA RPS12 has two conserved 3′-terminal adenines that direct a dominant non-canonical edit modulated by KREH2 (13). Thus, specific bases outside the classified canonical guiding region in some initiator guides may direct strong KREH2-modulated pausing in early editing. In the second step of the concerted model, a non-cognate guide gCOX3 [mO_235(Ia)_gCOX3(206-250)] installs the remaining 3′ HFE sequence. Importantly, +2U at site 33 would create a binding site for gCOX3 in ~60% of the ND7 transcriptome. About half of these molecules (~30% of ND7 3′ reads) contained the consensus 3′ HFE in samples without a knockdown. Thus, KREH2-promoted concerted non-canonical guiding, involving "unexpected" edits by a cognate guide (gRNA-1) and "expected" edits by a noncognate guide (gCOX3), install the consensus 3′ HFE. Besides sites 33 and 37-43 in the 3′ HFE, sites 49 and 55 and other sites examined exhibited large KREH2-promoted NC/C ratios (near 1 or >1; **Supplementary Table ST3**). Thus, substantial KREH2-dependent pausing occurs at the corresponding IPS48 and IPS54, among other sites along ND7 3′. However, the PCF-specific major checkpoint in ND7 3′ involves KREH2-promoted 3′ HFE formation and pausing in early editing progression.

Analyses of all amplicons in samples confirmed the absence of canonical editing upstream of the 3′ HFE. The KREH2-modulated non-cognate guide gCOX3 derails canonical editing beyond block 1, so it is a terminator. Short non-canonical junctions of variable composition and unclear sources usually followed the 3′ HFE. Interestingly, the junction length (JL) increased in the KREH2 knockdown (**Fig. 7**; **Supplementary Figures S9-S11)**. Reported knockdowns of several RESC proteins increased the JL immediately upstream of canonical editing in mRNA transcripts (35,70,71), suggesting that this may be a general feature of loss-of-function of editing factors examined so far.

Overall, a PCF-specific major checkpoint in ND7 3′ in early editing involves KREH2-dependent selection of housekeeping initiator gRNA-1 and novel terminator gRNA gCOX3. Surprisingly, PCF cells exploit the highly efficient initiator guide to install the most frequent non-canonical edit in ND7 3′, which also causes robust pausing in editing progression. Thus, KREH2-dependent differential gRNA selection establishes negative control at an early checkpoint in ND7 3′ editing in PCF (**Fig. 10A**).

KREH2-dependent gRNA selection may occur throughout the editing domain in ND7 3′

While ~30% of the ND7 3′ transcriptome in PCF carried the repressive 3′ HFE, ~10% and 20% of ND7 3′ transcripts in mtRNA or native RESC6 complexes, respectively, lacked the 3′ HFE and thus exhibited canonical editing beyond block 1 (**Fig. 5F; Fig. 6A**). Notably, KREH2 knockdown in PCF significantly increased canonical editing and decreased non-canonical editing at most sites examined in ND7 3′ without the 3′ HFE. This reversed control upon KREH2 knockdown was more robust in RESC6 complexes than in total mtRNA, suggesting active modulation by REH2C in holo-editosomes. These observations have two important implications for our proposed model of ND7 3′ editing repression in PCF. First, enhanced maturation of full editing blocks in the knockdown indicated that KREH2 controls editing by an entire gRNA guiding domain and not just of specific bases in the guiding domain. Second, KREH2-induced editing pausing and possibly termination may affect most sites along ND7 3′. So, KREH2-dependent guide usage, cognate and non-cognate, may occur throughout ND7 3′, not only at the major early checkpoint (**Fig. 10B**). KREH2 may also modulate editing guided by specific gRNA bases outside the classified guiding regions in other guides, as shown for conserved terminal adenines in initiator gRNAs in ND7 3′ in this study, and in RPS12 in a prior study (13). Although KREH2 may control gRNA usage or even usage of specific bases in a gRNA, mRNA turnover control may also be important during development (72). The frequency of gRNAs, cognate or non-cognate, or mRNA-gRNA duplex thermodynamic stability (predicted ΔG) may not be critical determinants in developmental ND7 3′ control. Thus, KREH2-dependent gRNA selection of cognate and non-cognate guides may repress editing throughout ND7 3′ in PCF, but mostly at an early checkpoint (**Fig. 10**).

RNA conformation in the ND7 3′ early checkpoint may hinder potential "repair" editing in PCF

The 3′ HFE appears to abolish further editing action in ND7 3′. However, cognate gRNA-2 could potentially bind downstream of the 3′ HFE and result in editing that replaces this repressive element with canonical sequence. Although "repair" editing seems feasible, the 3′ HFE exhibits a massive accumulation in PCF. To solve this conundrum, we propose that 3′ HFE removal may be blocked by at least two colluding processes in PCF. First, as discussed above, KREH2 downregulates gRNA-2 function (**Fig. 4E; Fig. 6A**). Second, 3′ HFE may form a stable RNA fold determined by DMS-MaPseq blocking mRNA 3′ terminus access to gRNA-2 (**Fig. 10**). Our experimentally determined RNA fold *in vitro*, suggests that the 3′ HFE engages

in an extended co-axial helix potentially hijacking the gRNA-2 anchor binding site. Generation of the proposed repressive RNA fold (requiring canonical block 1 plus 3′ HFE formation) would consume ~18% of the ATP needed for full-editing maturation of the entire ND7 3′ domain, effectively saving energy by blocking all editing action early in the process in PCF. It will be important to examine the structure of mRNA folded in native conditions as RNA secondary structure may be influenced by protein binding, and environmental conditions during development.

Decreased total editing action upon KREH2 knockdown contributes to the observed reduced accumulation of mature ND7 3′ in PCF and BSF

KREH2 knockdown decreased non-canonical editing and increased canonical editing at most sites examined in ND7 3′ in PCF examined by RNAseq. However, KREH2 knockdown decreases the accumulation of fully edited mature mRNAs in both PCF and BSF, as measured by RT-qPCR (38). RT-qPCR typically quantitates full editing at a 5′ block in editing substrates. This apparent conundrum may be explained by a secondary effect of KREH2 knockdown resulting in a general decrease in total editing action (**Fig. 3**). Similarly, the knockdown of core editing proteins in RECC and RECC reduces the accumulation of fully edited transcripts in PCF and BSF (5,73). The cumulative loss in total edits (canonical and non-canonical) at most editing sites likely causes a net loss in full editing in these knockdowns. A reduction in general editing action may be partly due to disruptions in editing machinery assembly. For example, the KREH2 knockdown shifted the sedimentation of RESC proteins (43).

ND7 3′ maturation hindrance via KREH2-dependent opposite modulation of canonical and non-canonical editing and putative repressive structure is specific to PCF. This negative control system is somehow reversed in the BSF lifecycle stage, where KREH2 promotes ND7 3′ maturation. Mechanisms driving a functional KREH2 switch between negative and positive control modes during development could be protein-dependent or independent. In the REH2C editing complex, KH2F1 and KH2F2 associate with KREH2 in an RNA-free manner *in vivo* or in recombinant form (**Fig. 1B**) (32,38,39). Temperature could also play a role in editing control, as it directly influences RNA structure dynamics (22,74). "RNA thermometers", perhaps aided by KREH2, may regulate editing in response to shifts in temperature and intracellular conditions. Thermometers in trypanosomes, as in other

systems, including in plants and chloroplast in green alga (74,75), could influence mRNA and or gRNA secondary structure, or RNA helicase conformation, masking or promoting productive RNA hybrids, cognate or non-cognate. Interestingly, temperature reduction promoted editing efficiency of some cytochrome mRNAs. In the case of COX2, expression of the accessory protein p22 and depletion of RDK1, a kinase that inhibits BSF to PCF differentiation, was coupled to cold-induced COX2 editing (64).

Implications of gRNA-directed non-canonical editing control by REH2C on the constructive neutral evolution hypothesis and a multifunctional gRNA transcriptome

The identification of gRNA-programmed non-canonical editing that is regulated by specific factors challenges the view of RNA editing as an example of constructive neutral evolution (CNE). The CNE model postulates that molecular mechanisms, including RNA editing, can evolve without evolutionary benefits (76,77). Under this view, non-canonical editing junctions of unpredictable composition and unclear origin occur without positive selection and are neutrally fixed i.e., by genetic drift. A regulated 3′ HFE in ND7 3′ and in mRNA A6 in a prior study (38) indicate that at least some non-canonical editing has evolved via natural selection. However, as previously suggested, the initial evolution and fixation of editing could have been neutral, but subsequent elaboration, including adaptation, may have fixed new functions that could be beneficial. At least some editing could have evolved via CNE, but it then represented a new playground for Darwinian-type evolution to take place.

Opposite modulation by KREH2 on noncognate and cognate guiding along the ND7 3′ domain suggests that at least part of the gRNA transcriptome is potentially multifunctional. The regulatory proteins that drive developmental editing regulation have been elusive. We propose that the regulatory factors are present in the editing holoenzyme, including the helicase complex REH2C. We have shown that at least KREH2 modulates the novel gCOX3 and gCR4 guides, which, respectively, promote editing in their cognate mRNAs but install a repressive 3′ HFE in ~30% of non-cognate ND7 3′ and A6 transcriptomes, respectively (38). These and other possible KREH2-dependent regulatory gRNAs are potentially bifunctional. KREH2-dependent selection of gRNAs as positive (housekeeping) or negative (repressor) factors may be a general strategy to establish the level of mature edited mRNA. Additional studies are required to better understand KREH2-dependent gRNA selection control during

development. Regardless, it is increasingly clear that KREH2 purposely regulates non-canonical editing on a large-scale during trypanosome development (38). Based on the basic editing reaction alone, involving mRNA cleavage, U indels, and ATP-consuming RNA ligation (68,78,79), maturation of the entire ND7 3′ domain would consume ~18 times more ATP than just installing the ≥54 nt extended element, which includes the 3′ HFE (**Fig. 5D**). As proposed for A6, a general model invoking gRNA-programmed non-canonical editing to control canonical editing and, thus, energy efficiency would be a novel feature of trypanosomal biology. In ND7 3′, massive gRNA-1-programmed editing, including non-canonical, may titrate factors in a 'sponge effect,' limiting editing-based strategies to bypass early termination. KREH2-dependent concerted action by the initiator gRNA-1 and novel terminator gCOX3 near the mRNA 3′ terminus may improve general energy consumption in PCF where ND7 3′ maturation may be dispensable. A prior study proposed two terminator gRNAs for editing of mRNA COX3, which may introduce a short element derailing canonical editing (46). These elements installed by as yet unidentified guides were more abundant in BSF than in PCF and often included upstream junctions of variable composition (46). Whether the element sequence directed by these guides is structural or regulated by specific proteins was not addressed. In the kinetoplastid *L. pyrrohocoris*, some gRNAs were proposed to target non-cognate mRNAs. However, the examples reported might alter the ORF without derailing upstream editing (80). Our recent study in mRNA A6 identified a KREH2-modulated early checkpoint, where RNA structure formed by a 3′ HFE may occlude the mRNA 3′ terminus, preventing potential "repair" editing by the initiator gRNA-1 (38). The novel gCR4 gRNA that installs this 3′ HFE is under negative control in PCF, i.e., REH2 knockdown upregulates gCR4 usage, thereby 3′ HFE formation in A6, exclusively in procyclics. This gCR4 gRNA binds pre-edited block 1 sequence and abolishes all A6 canonical editing, potentially as an "anti-initiator" in the targeted transcripts.

PCF-specific KREH2-dependent hindrance of ND7 3′ editing maturation is consistent with long-established preferential editing maturation of BSF mRNAs for mitochondrially encoded subunits of complex I "cI". Prior studies reported that differentiation into transmission-competent BSF parasites appears associated with up-regulation of cI (20,21) and characterization of cI subunits in BSF showed the presence of multi-subunit complexes, but as to whether a complete cI is assembled remains unclear (18). Surprisingly, the deletion of

two cI subunits suggested that electron transfer within cI is not essential in the slender BSF stage and that cI does not contribute significantly to NADH dehydrogenase activity in this stage. However, that study showed an exacerbated growth phenotype upon NDH2 ablation when cI was knocked out simultaneously. That suggests that cI is functional in slender forms, but there's redundancy with NDH2 (18). In PCF, cI was detected but may not participate in the electron transport chain (81). Thus, cI complexes may have diverged in the long evolutionary history of trypanosomes, and their precise functional role during development is unclear. Nevertheless, there is natural selection for efficient ND7 3′ canonical editing in BSF. ND7 3′ provides a robust model system of PCF-specific KREH2-dependent editing repression.

Thus, KREH2 is a key regulatory factor in stage-specific RNA editing in *T. brucei* development. The current study defined novel concepts important to understanding KREH2-dependent developmental regulation: (i) global control, positive and negative, of gRNA function by KREH2 and presumably other proteins in the REH2C complex, (ii) KREH2-dependent gRNA selection modulates editing fidelity (i.e., non-canonical vs. canonical edits) and thus the level of mature mRNA, (iii) major early checkpoints may involve the converse regulation of repressor guides, which install a structural element, and canonical housekeeping guides (model in **Fig. 10**). KREH2-regulated gRNAs with alternative functions in cognate and noncognate mRNAs suggests that the trypanosomal gRNA transcriptome may be multifunctional. Additional studies will determine whether REH2C proteins developmentally regulate editing in other mRNAs and whether major early checkpoints are common in both stages. Our findings in ND7 3′, together with our prior A6 report (38), challenge the popular hypothesis of U-indel RNA editing as an example of constructive editing evolution.

DATA AVAILABILITY

RNA sequencing data are deposited in the NCBI SRA under Bioproject ID PRJNA1004327.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank members of the Institute for Genome Sciences and Society at Texas A&M University for their excellent assistance with deep sequencing. The Afasizhev lab kindly provided the antiserum against RESC1/2 (GAP2/1) for RNA Immunoprecipitations. The Read lab kindly provided the antisera against RESC13 (KRGG2). Zihao Chen kindly assisted with microscopy. We are very grateful to Thavy Khem, Melanie Kuwahara, and Neida Murillo for their assistance in preparing Excel files and charts.

FUNDING

This work was supported by the National Science Foundation [1616845 to JCR., 2140153 to SM]; and Texas A&M X-grant [248543 to JCR]. Funding for open access charge: National Science Foundation [1616845].

CONFLICT OF INTEREST

NA

REFERENCES

1. Kostygov, A.Y., Karnkowska, A., Votypka, J., Tashyreva, D., Maciszewski, K., Yurchenko, V. and Lukes, J. (2021) Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol*, **11**, 200407.
2. Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gurtler, R.E., McKerrow, J., Reed, S. and Tarleton, R. (2008) Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of Clinical Investigation*, **118**, 1301-1310.

3. Zimmer, S.L. (2019) Revisiting Trypanosome Mitochondrial Genome Mysteries: Broader and Deeper. *Trends Parasitol*, **35**, 102-104.
4. Cruz-Reyes, J., Mooers, B.H.M., Doharey, P.K., Meehan, J. and Gulati, S. (2018) Dynamic RNA holo-editosomes with subcomplex variants: Insights into the control of trypanosome editing. *Wiley Interdiscip Rev RNA*, **9**, e1502.
5. Aphasizheva, I., Alfonzo, J., Carnes, J., Cestari, I., Cruz-Reyes, J., Goringer, H.U., Hajduk, S., Lukes, J., Madison-Antenucci, S., Maslov, D.A. *et al.* (2020) Lexis and Grammar of Mitochondrial RNA Processing in Trypanosomes. *Trends Parasitol*, **36**, 337-355.
6. Jensen, R.E. and Englund, P.T. (2012) Network news: the replication of kinetoplast DNA. *Annu Rev Microbiol*, **66**, 473-491.
7. Lukes, J., Guilbride, D.L., Votypka, J., Zikova, A., Benne, R. and Englund, P.T. (2002) Kinetoplast DNA network: evolution of an improbable structure. *Eukaryotic Cell*, **1**, 495-502.
8. Cooper, S., Wadsworth, E.S., Schnaufer, A. and Savill, N.J. (2022) Organization of minicircle cassettes and guide RNA genes in Trypanosoma brucei. *RNA*, **28**, 972-992.
9. Cooper, S., Wadsworth, E.S., Ochsenreiter, T., Ivens, A., Savill, N.J. and Schnaufer, A. (2019) Assembly and annotation of the mitochondrial minicircle genome of a differentiation-competent strain of Trypanosoma brucei. *Nucleic Acids Res*, **47**, 11304-11325.
10. Kirby, L.E., Sun, Y., Judah, D., Nowak, S. and Koslowsky, D. (2016) Analysis of the Trypanosoma brucei EATRO 164 Bloodstream Guide RNA Transcriptome. *PLoS Negl Trop Dis*, **10**, e0004793.
11. Carnes, J., McDermott, S., Anupama, A., Oliver, B.G., Sather, D.N. and Stuart, K. (2017) In vivo cleavage specificity of Trypanosoma brucei editosome endonucleases. *Nucleic Acids Research*, **45**, 4667-4686.
12. Simpson, R.M., Bruno, A.E., Bard, J.E., Buck, M.J. and Read, L.K. (2016) High-throughput sequencing of partially edited trypanosome mRNAs reveals barriers to editing progression and evidence for alternative editing. *RNA*, **22**, 677-695.
13. Kumar, V., Ivens, A., Goodall, Z., Meehan, J., Doharey, P.K., Hillhouse, A., Hurtado, D.O., Cai, J.J., Zhang, X., Schnaufer, A. *et al.* (2020) Site-specific and mRNA-specific control of accurate mRNA editing by a helicase complex in trypanosomes. *RNA*.

14. Stuart, K., Allen, T.E., Heidmann, S. and Seiwert, S.D. (1997) RNA editing in kinetoplastid protozoa. *Microbiology and Molecular Biology Reviews : MMBR*, **61**, 105-120.

15. Read, L.K., Lukes, J. and Hashimi, H. (2016) Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdiscip Rev RNA*, **7**, 33-51.

16. Zimmer, S.L., Simpson, R.M. and Read, L.K. (2018) High throughput sequencing revolution reveals conserved fundamentals of U-indel editing. *Wiley Interdiscip Rev RNA*, e1487.

17. Vickerman, K. (1985) Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin*, **41**, 105-114.

18. Surve, S.V., Jensen, B.C., Heestand, M., Mazet, M., Smith, T.K., Bringaud, F., Parsons, M. and Schnaufer, A. (2017) NADH dehydrogenase of Trypanosoma brucei is important for efficient acetate production in bloodstream forms. *Molecular and Biochemical Parasitology*, **211**, 57-61.

19. Zikova, A., Verner, Z., Nenarokova, A., Michels, P.A.M. and Lukes, J. (2017) A paradigm shift: The mitoproteomes of procyclic and bloodstream Trypanosoma brucei are comparably complex. *PLoS Pathogens*, **13**, e1006679.

20. Gunasekera, K., Wuthrich, D., Braga-Lagache, S., Heller, M. and Ochsenreiter, T. (2012) Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. *BMC Genomics*, **13**, 556.

21. Priest, J.W. and Hajduk, S.L. (1994) Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei. *Journal of Bioenergetics and Biomembranes*, **26**, 179-191.

22. Koslowsky, D.J., Sun, Y., Hindenach, J., Theisen, T. and Lucas, J. (2013) The insect-phase gRNA transcriptome in Trypanosoma brucei. *Nucleic Acids Research*.

23. Rusche, L.N., Cruz-Reyes, J., Piller, K.J. and Sollner-Webb, B. (1997) Purification of a functional enzymatic editing complex from Trypanosoma brucei mitochondria. *EMBO J*, **16**, 4069-4081.

24. Stuart, K., Panigrahi, A.K., Schnaufer, A., Drozdz, M., Clayton, C. and Salavati, R. (2002) Composition of the editing complex of Trypanosoma brucei. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **357**, 71-79.

25. Aphasizhev, R., Aphasizheva, I., Nelson, R.E., Gao, G., Simpson, A.M., Kang, X., Falick, A.M., Sbicego, S. and Simpson, L. (2003) Isolation of a U-insertion/deletion editing complex from Leishmania tarentolae mitochondria. *EMBO J*, **22**, 913-924.
26. Carnes, J., Trotter, J.R., Peltan, A., Fleck, M. and Stuart, K. (2008) RNA editing in Trypanosoma brucei requires three different editosomes. *Mol Cell Biol*, **28**, 122-130.
27. Panigrahi, A.K., Zikova, A., Dalley, R.A., Acestor, N., Ogata, Y., Anupama, A., Myler, P.J. and Stuart, K.D. (2008) Mitochondrial complexes in Trypanosoma brucei: a novel complex and a unique oxidoreductase complex. *Molecular & Cellular Proteomics : MCP*, **7**, 534-545.
28. Weng, J., Aphasizheva, I., Etheridge, R.D., Huang, L., Wang, X., Falick, A.M. and Aphasizhev, R. (2008) Guide RNA-binding complex from mitochondria of trypanosomatids. *Mol Cell*, **32**, 198-209.
29. Hashimi, H., Zikova, A., Panigrahi, A.K., Stuart, K.D. and Lukes, J. (2008) TbRGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex. *RNA*, **14**, 970-980.
30. Ammerman, M.L., Downey, K.M., Hashimi, H., Fisk, J.C., Tomasello, D.L., Faktorova, D., Kafkova, L., King, T., Lukes, J. and Read, L.K. (2012) Architecture of the trypanosome RNA editing accessory complex, MRB1. *Nucleic Acids Research*, **40**, 5637-5650.
31. Madina, B.R., Kumar, V., Metz, R., Mooers, B.H., Bundschuh, R. and Cruz-Reyes, J. (2014) Native mitochondrial RNA-binding complexes in kinetoplastid RNA editing differ in guide RNA composition. *RNA*, **20**, 1142-1152.
32. Kumar, V., Madina, B.R., Gulati, S., Vashisht, A.A., Kanyumbu, C., Pieters, B., Shakir, A., Wohlschlegel, J.A., Read, L.K., Mooers, B.H. *et al.* (2016) REH2C Helicase and GRBC Subcomplexes May Base Pair through mRNA and Small Guide RNA in Kinetoplastid Editosomes. *J Biol Chem*, **291**, 5753-5764.
33. Aphasizheva, I., Zhang, L., Wang, X., Kaake, R.M., Huang, L., Monti, S. and Aphasizhev, R. (2014) RNA binding and core complexes constitute the U-insertion/deletion editosome. *Mol Cell Biol*, **34**, 4329-4342.
34. Dixit, S., Muller-McNicoll, M., David, V., Zarnack, K., Ule, J., Hashimi, H. and Lukes, J. (2017) Differential Binding of Mitochondrial Transcripts by MRB8170 and MRB4160 Regulates Distinct Editing Fates of Mitochondrial mRNA in Trypanosomes. *MBio*, **8(1)**, e02288-02216.

35. McAdams, N.M., Simpson, R.M., Chen, R., Sun, Y. and Read, L.K. (2018) MRB7260 is essential for productive protein-RNA interactions within the RNA editing substrate binding complex during trypanosome RNA editing. *RNA*, **24**, 540-556.

36. Liu, S., Wang, H., Li, X., Zhang, F., Lee, J.K.J., Li, Z., Yu, C., Hu, J.J., Zhao, X., Suematsu, T. *et al.* (2023) Structural basis of gRNA stabilization and mRNA recognition in trypanosomal RNA editing. *Science*, **381**, eadg4725.

37. Dolce, L.G., Nesterenko, Y., Walther, L., Weis, F. and Kowalinski, E. (2023) Structural basis for guide RNA selection by the RESC1-RESC2 complex. *Nucleic Acids Research*, **51**, 4602-4612.

38. Meehan, J., McDermott, S.M., Ivens, A., Goodall, Z., Chen, Z., Yu, Z., Woo, J., Rodshagen, T., McCleskey, L., Sechrist, R. *et al.* (2023) Trypanosome RNA helicase KREH2 differentially controls non-canonical editing and putative repressive structure via a novel proposed 'bifunctional' gRNA in mRNA A6. *Nucleic Acids Research*, **51**, 6944-6965.

39. Kumar, V., Doharey, P.K., Gulati, S., Meehan, J., Martinez, M.G., Hughes, K., Mooers, B.H.M. and Cruz-Reyes, J. (2019) Protein features for assembly of the RNA editing helicase 2 subcomplex (REH2C) in Trypanosome holo-editosomes. *PloS One*, **14**, e0211525.

40. Dubey, A.P., Tylec, B.L., Mishra, A., Sortino, K., Chen, R., Sun, Y. and Read, L.K. (2023) KREH1 RNA helicase activity promotes utilization of initiator gRNAs across multiple mRNAs in trypanosome RNA editing. *Nucleic Acids Research*, **51**, 5791-5809.

41. Hernandez, A., Madina, B.R., Ro, K., Wohlschlegel, J.A., Willard, B., Kinter, M.T. and Cruz-Reyes, J. (2010) REH2 RNA helicase in kinetoplastid mitochondria: ribonucleoprotein complexes and essential motifs for unwinding and guide RNA (gRNA) binding. *J Biol Chem*, **285**, 1220-1228.

42. Madina, B.R., Kumar, V., Mooers, B.H. and Cruz-Reyes, J. (2015) Native Variants of the MRB1 Complex Exhibit Specialized Functions in Kinetoplastid RNA Editing. *PloS One*, **10**, e0123441.

43. Hashimi, H., Cicova, Z., Novotna, L., Wen, Y.Z. and Lukes, J. (2009) Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. *RNA*, **15**, 588-599.

44. Koslowsky, D.J., Bhat, G.J., Perrollaz, A.L., Feagin, J.E. and Stuart, K. (1990) The MURF3 gene of T. brucei contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. *Cell*, **62**, 901-911.

45. Koslowsky, D.J., Riley, G.R., Feagin, J.E. and Stuart, K. (1992) Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of Trypanosoma brucei. *Mol Cell Biol*, **12**, 2043-2049.

46. Smith, J.T., Jr., Dolezelova, E., Tylec, B., Bard, J.E., Chen, R., Sun, Y., Zikova, A. and Read, L.K. (2020) Developmental regulation of edited CYb and COIII mitochondrial mRNAs is achieved by distinct mechanisms in Trypanosoma brucei. *Nucleic Acids Research*, **48**, 8704-8723.

47. McDermott, S.M., Guo, X., Carnes, J. and Stuart, K. (2015) Differential Editosome Protein Function between Life Cycle Stages of Trypanosoma brucei. *J Biol Chem*, **290**, 24914-24931.

48. McDermott, S.M., Carnes, J. and Stuart, K. (2019) Editosome RNase III domain interactions are essential for editing and differ between life cycle stages in Trypanosoma brucei. *RNA*, **25**, 1150-1163.

49. Dean, S., Gould, M.K., Dewar, C.E. and Schnaufer, A.C. (2013) Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes. *Proc Natl Acad Sci U S A*, **110**, 14741-14746.

50. Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B. *et al.* (2007) Functional genomics in Trypanosoma brucei: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Molecular and Biochemical Parasitology*, **154**, 103-109.

51. McDermott, S.M., Pham, V., Lewis, I., Tracy, M. and Stuart, K. (2023) mt-LAF3 is a pseudouridine synthase ortholog required for mitochondrial rRNA and mRNA gene expression in Trypanosoma brucei. *International Journal for Parasitology*, **53**, 573-583.

52. Panigrahi, A.K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R.A., Acestor, N., Myler, P.J. and Stuart, K.D. (2009) A comprehensive analysis of Trypanosoma brucei mitochondrial proteome. *Proteomics*, **9**, 434-450.

53. Breidbach, T., Ngazoa, E. and Steverding, D. (2002) Trypanosoma brucei: in vitro slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms. *Exp Parasitol*, **101**, 223-230.

54. Carnes, J., Trotter, J.R., Ernst, N.L., Steinberg, A. and Stuart, K. (2005) An essential RNase III insertion editing endonuclease in Trypanosoma brucei. *Proc Natl Acad Sci U S A*, **102**, 16614-16619.

55. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402-408.

56. Brenndorfer, M. and Boshart, M. (2010) Selection of reference genes for mRNA quantification in Trypanosoma brucei. *Molecular and Biochemical Parasitology*, **172**, 52-55.

57. Carnes, J., Gendrin, C., McDermott, S.M. and Stuart, K. (2023) KRGG1 function in RNA editing in Trypanosoma brucei. *RNA*, **29**, 228-240.

58. Riemondy, K.A., Sheridan, R.M., Gillen, A., Yu, Y., Bennett, C.G. and Hesselberth, J.R. (2017) valr: Reproducible genome interval analysis in R. *F1000Res*, **6**, 1025.

59. Tomezsko, P.J., Corbin, V.D.A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M.D., McIntosh, L., Papenfuss, A.T., Emery, A. *et al.* (2020) Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature*, **582**, 438-442.

60. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.

61. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974-1975.

62. Wong, R.G., Kazane, K., Maslov, D.A., Rogers, K., Aphasizhev, R. and Simpson, L. (2015) U-insertion/deletion RNA editing multiprotein complexes and mitochondrial ribosomes in Leishmania tarentolae are located in antipodal nodes adjacent to the kinetoplast DNA. *Mitochondrion*, **25**, 76-86.

63. Ammerman, M.L., Hashimi, H., Novotna, L., Cicova, Z., McEvoy, S.M., Lukes, J. and Read, L.K. (2011) MRB3010 is a core component of the MRB1 complex that facilitates an early step of the kinetoplastid RNA editing process. *RNA*, **17**, 865-877.

64. Smith, J.T., Jr., Tylec, B., Naguleswaran, A., Roditi, I. and Read, L.K. (2023) Developmental dynamics of mitochondrial mRNA abundance and editing reveal roles for temperature and the differentiation-repressive kinase RDK1 in cytochrome oxidase subunit II mRNA editing. *mBio*, **14**, e0185423.

65. Sabatini, R. and Hajduk, S.L. (1995) RNA ligase and its involvement in guide RNA/mRNA chimera formation. Evidence for a cleavage-ligation mechanism of Trypanosoma brucei mRNA editing. *J Biol Chem*, **270**, 7233-7240.

66. Schmid, B., Read, L.K., Stuart, K. and Goringer, H.U. (1996) Experimental verification of the secondary structures of guide RNA-pre-mRNA chimaeric molecules in Trypanosoma brucei. *European Journal of Biochemistry / FEBS*, **240**, 721-731.

67. Seiwert, S.D. and Stuart, K. (1994) RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro. *Science*, **266**, 114-117.

68. Seiwert, S.D., Heidmann, S. and Stuart, K. (1996) Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. *Cell*, **84**, 831-841.

69. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S. and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods*, **14**, 75-82.

70. Tylec, B.L., Simpson, R.M., Kirby, L.E., Chen, R., Sun, Y., Koslowsky, D.J. and Read, L.K. (2019) Intrinsic and regulated properties of minimally edited trypanosome mRNAs. *Nucleic Acids Research*, **47**, 3640-3657.

71. McAdams, N.M., Harrison, G.L., Tylec, B.L., Ammerman, M.L., Chen, R., Sun, Y. and Read, L.K. (2019) MRB10130 is a RESC assembly factor that promotes kinetoplastid RNA editing initiation and progression. *RNA*, **25**, 1177-1191.

72. Aphasizheva, I. and Aphasizhev, R. (2021) Mitochondrial RNA quality control in trypanosomes. *Wiley Interdiscip Rev RNA*, **12**, e1638.

73. Cruz-Reyes J., M.B.H.M., Kumar K., Doharey P.K., Meehan J., Chaparro L. (2018) In J, C.-R. and M, G. (eds.), *RNA Metabolism in Mitochondria*. Springer, Cham, Vol. 34.

74. Chung, K.P., Loiacono, F.V., Neupert, J., Wu, M. and Bock, R. (2023) An RNA thermometer in the chloroplast genome of Chlamydomonas facilitates temperature-controlled gene expression. *Nucleic Acids Research*, **51**, 11386-11400.

75. Thomas, S.E., Balcerowicz, M. and Chung, B.Y. (2022) RNA structure mediated thermoregulation: What can we learn from plants? *Front Plant Sci*, **13**, 938570.

76. Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *Journal of Molecular Evolution*, **49**, 169-181.

77. Gray, M.W., Lukes, J., Archibald, J.M., Keeling, P.J. and Doolittle, W.F. (2010) Cell biology. Irremediable complexity? *Science*, **330**, 920-921.

78. Kable, M.L., Seiwert, S.D., Heidmann, S. and Stuart, K. (1996) RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science*, **273**, 1189-1195.

79. Cruz-Reyes, J. and Sollner-Webb, B. (1996) Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. *Proc Natl Acad Sci U S A*, **93**, 8901-8906.

80. Gerasimov, E.S., Gasparyan, A.A., Afonin, D.A., Zimmer, S.L., Kraeva, N., Lukes, J., Yurchenko, V. and Kolesnikov, A. (2021) Complete minicircle genome of Leptomonas pyrrhocoris reveals sources of its non-canonical mitochondrial RNA editing events. *Nucleic Acids Research*, **49**, 3354-3370.

81. Verner, Z., Cermakova, P., Skodova, I., Kriegova, E., Horvath, A. and Lukes, J. (2011) Complex I (NADH:ubiquinone oxidoreductase) is active in but non-essential for procyclic Trypanosoma brucei. *Molecular and Biochemical Parasitology*, **175**, 196-200.

FIGURES

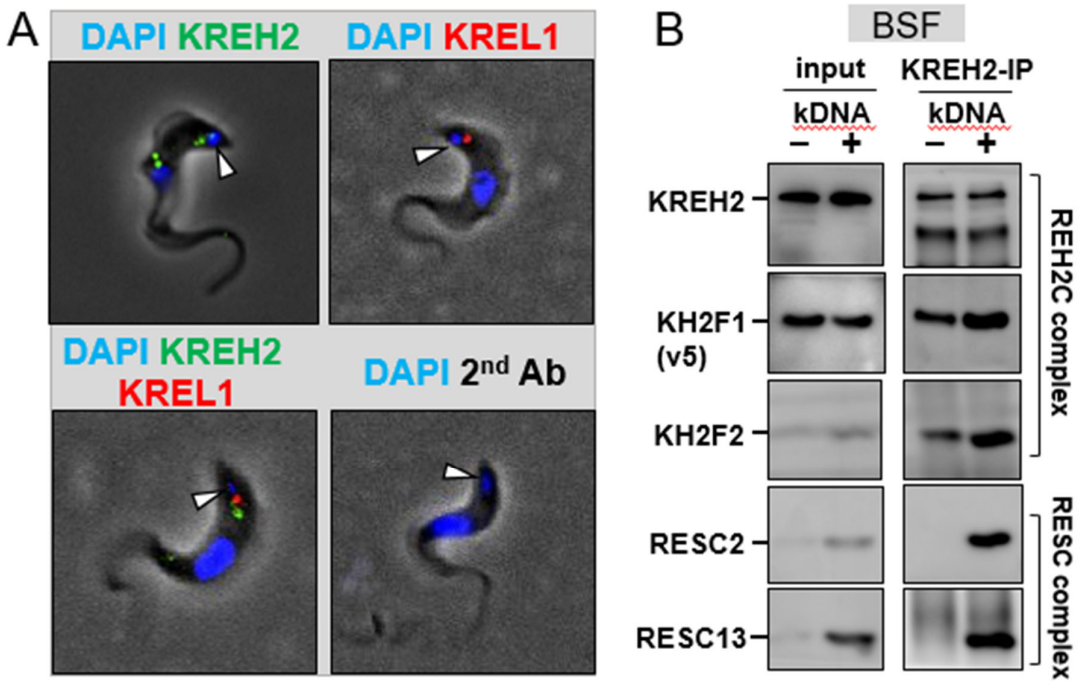


Figure 1. KREH2 localization and association with proteins in REH2C in BSF cells. (A) Immunofluorescent microscopy of BSF *T. brucei*. Cells were imaged for marker proteins in editing complexes: KREH2 (REH2C) and KREL1 (RECC). White arrows point to DAPI-stained kDNA in each cell. **(B)** Western blot of mt-extracts (input) and KREH2 immunoprecipitation (KREH2-IP) from a BSF dyskinetoplastic cell line that we made via acriflavine treatment as described in the methods section, which is devoid of kDNA (-kDNA). The parental cell line that was not treated with acriflavine and contains kDNA (+kDNA) was included as a control. These BSF cell lines are conditional nulls of KH2F1 that constitutively express KH2F1-v5 tagged protein, as described in the methods section. KREH2, KH2F1-v5, KH2F2 (all in REH2C), and RESC2 and RESC13 (in RESC) were examined.

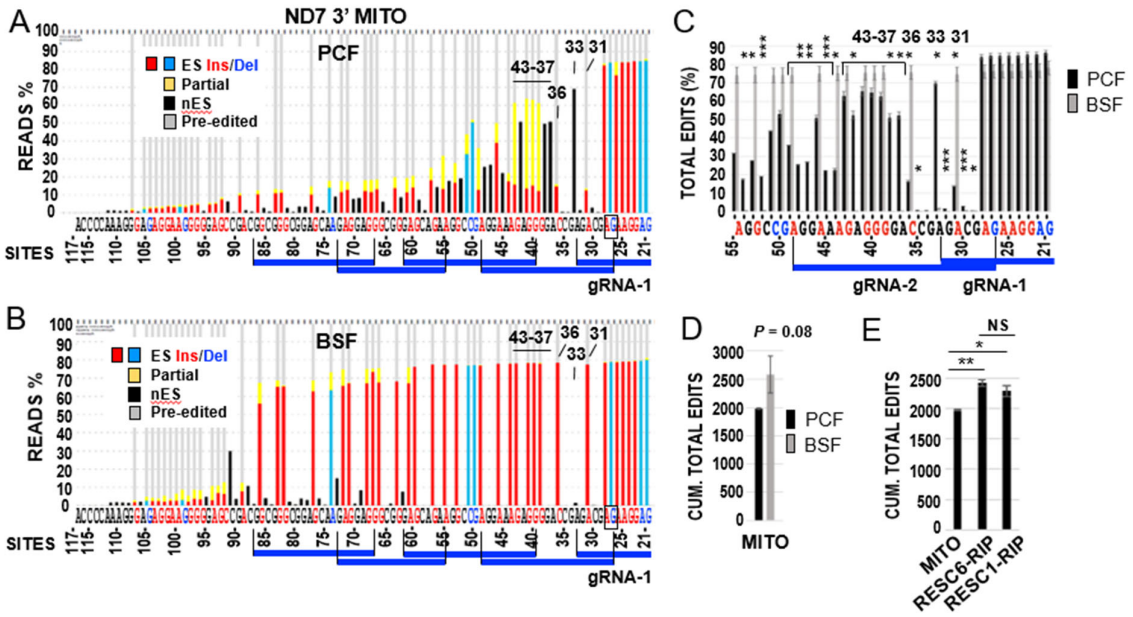


Figure 2. Comparison of ND7 3' total editing in PCF and BSF cells. 'Snapshots' of typical ND7 3' amplicon RNA-seq data sets from **(A)** PCF and **(B)** BSF. Stacked histograms show all possible editing events at each site in replicate mtRNA (Mito) samples from cells containing a KREH2-RNAi construct -Tet (i.e., uninduced), as in all other panels below, which used the same uninduced construct. Color-coded nucleotides are just 3' to canonical sites for U-insertion (Ins, red), U-deletion (Del, blue), or sites not expected to change in mature mRNA (black) (see **Supplementary Table S1**; glossary of terms). Bars represent the percentage of canonical insertion (red), deletion (blue), or non-canonical edits (yellow) at canonical sites or edits at sites not expected to change (black). Blocks of canonical editing thought to be directed by known gRNAs are indicated by indigo lines. **(C)** Site-by-site analysis of total edits across gRNA-1 and gRNA-2 (through site 55) in mtRNA from PCF vs. BSF cells. **(D)** Cumulative total edits in PCF vs BSF mtRNA. **(E)** Cumulative total edits from mtRNA, RESC6-RNA-Immunoprecipitation (RIP), and RESC1-RIP. The cumulative value at the most 5' site (site 117) in the amplicon was plotted. Full-amplicon analyses are in **Supplementary Fig. S2C**. As noted above, all cells in this figure are uninduced. Average and error bars of biological replicates to determine *P*-values *** < 0.005, ** < 0.05, * < 0.5 were annotated. See Material and Methods and **Supplementary Table S3** for additional details on statistical analysis.

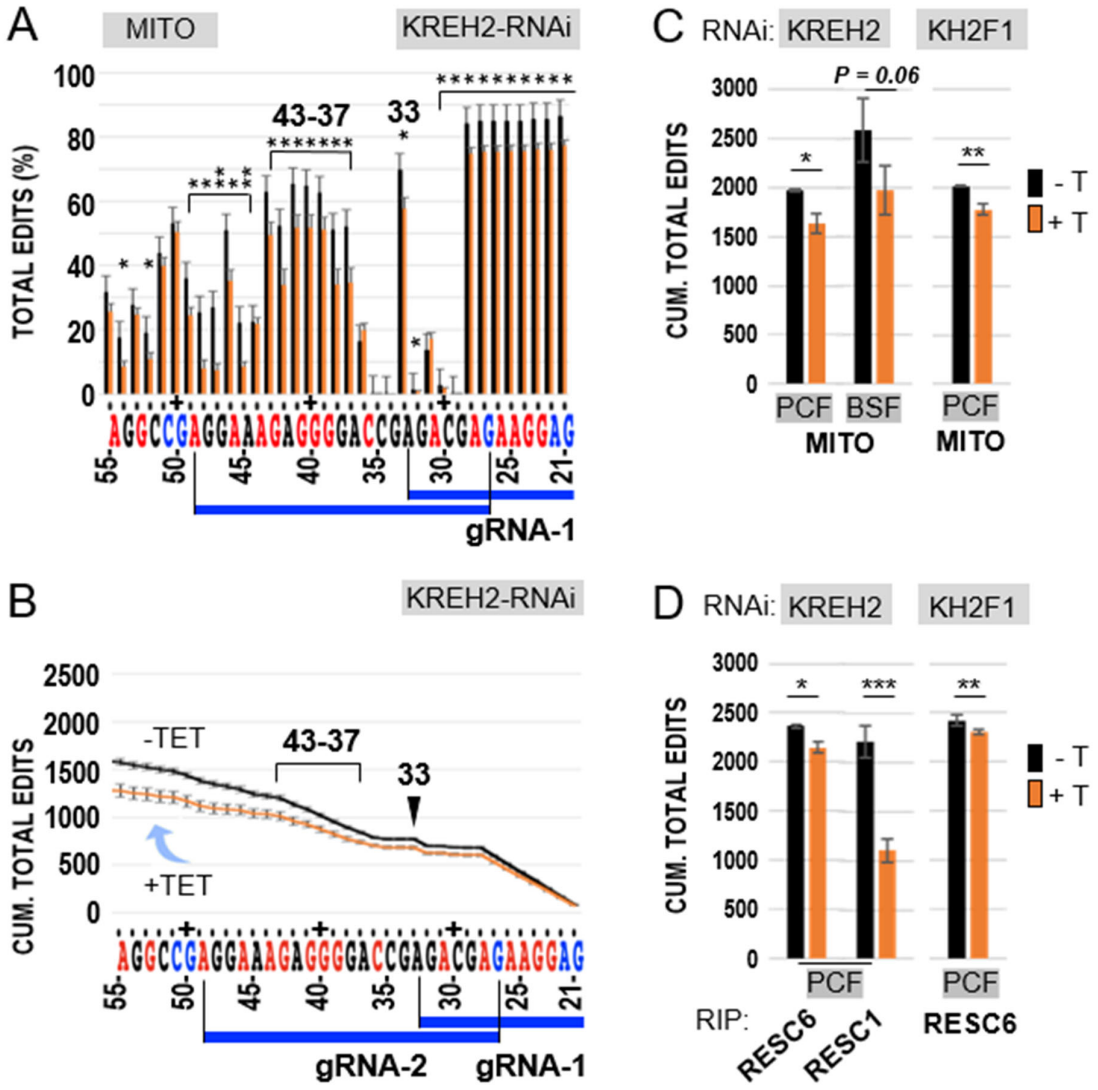


Figure 3. Analyses of ND7 3′ editing domain total editing upon KREH2-RNAi and KH2F1-RNAi in PCF and KREH2-RNAi in BSF. (A) Site-by-site, and **(B)** cumulative total edits across gRNA-1 and gRNA-2 (through site 55) in PCF mtRNA ± KREH2-RNAi. RNAi knockdowns decreased the targeted protein by ~80% (see methods section). Sites 33 and 37-43 of particular interest in ND7 3′ early editing, as described in the text, are highlighted. **(C)** Cumulative total edits up to the most 5′ site (site 117) in the ND7 3′ domain amplicon in PCF mtRNA ± KREH2-RNAi or KH2F1-RNAi and BSF mtRNA ± KREH2-RNAi. **(D)** Cumulative total edits as in panel C but in RESC6- and RESC1-RIPs in PCF KREH2-RNAi or KH2F1-RNAi. Full-amplicon analyses are in **Supplementary Fig. S3**. ± Tet (labeled ± T). Statistical analyses are as in **Fig. 2**.

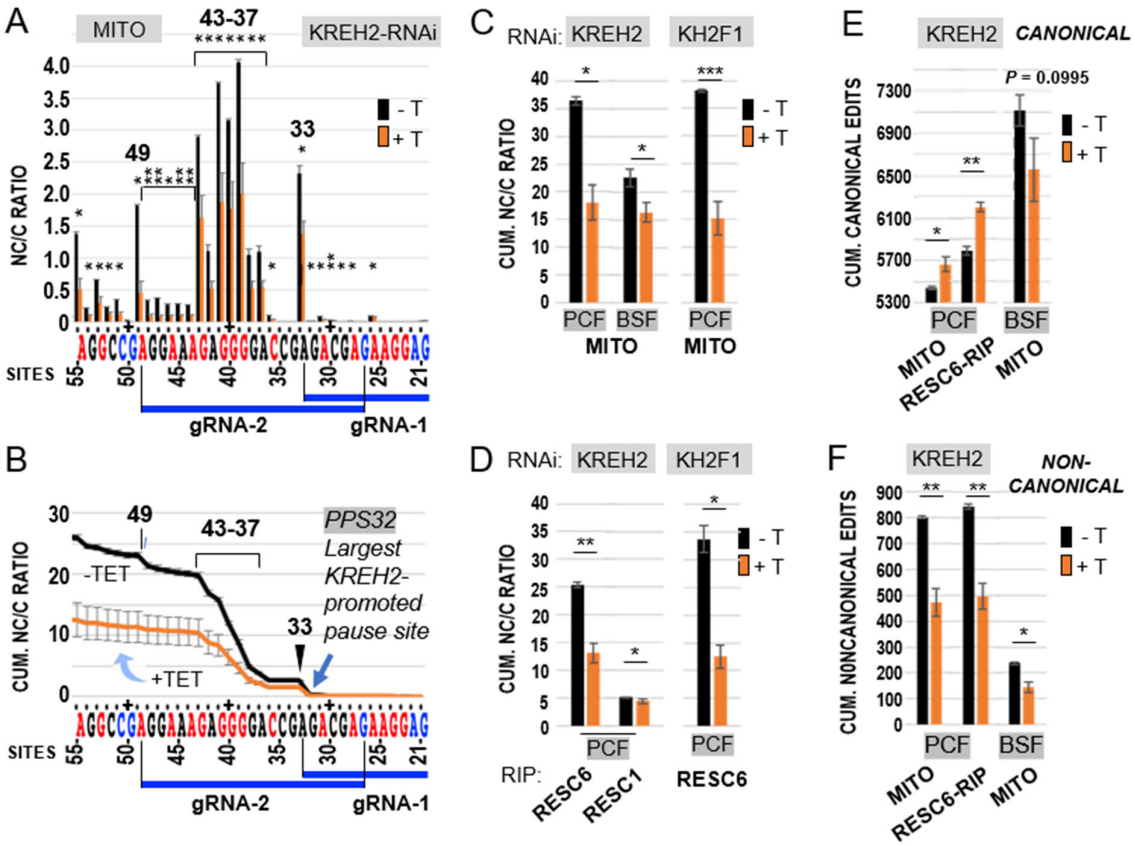


Figure 4. Analyses of ND7 3′ domain editing fidelity (NC/C) upon KREH2-RNAi or KH2F1-RNAi in mtRNA or RESC. (A) Site-by-site and **(B)** cumulative NC/C ratios across gRNA blocks 1-2 in PCF mtRNA ± KREH2-RNAi. NC/C ratios are the percentage of non-canonical reads divided by the percentage of canonical reads at the same site. Sites 33, 37-43, and 49 (highlighted) exhibited particularly high NC/C ratios in the replicates. **(C)** Cumulative NC/C ratios at the most 5′ site (site 117) in the ND7 3′ domain amplicon in PCF mtRNA ± KREH2-RNAi or KH2F1-RNAi and BSF mtRNA ± KREH2-RNAi. **(D)** Cumulative NC/C ratios as in panel C but in RESC6- and RESC1-RIPs in PCF ± KREH2-RNAi or KH2F1-RNAi. **(E)** Cumulative canonical and **(F)** cumulative non-canonical edits at the most 5′ site (site 117) in mtRNA and RESC6-RIP in PCF or BSF ± KREH2-RNAi. All RNAi is ± Tet (labeled ± T). Note that the +Tet decreased NC/C, i.e., increased editing fidelity. Full-amplicon analyses are in **Supplementary Fig. S4** and **Table ST3.** The highest NC/C in the amplicon at site 33 (arrowhead) caused the largest KREH2-promoted pause at site 32 (PPS32) in the sequence examined. Other annotations and statistical analyses are as in **Fig. 2**.

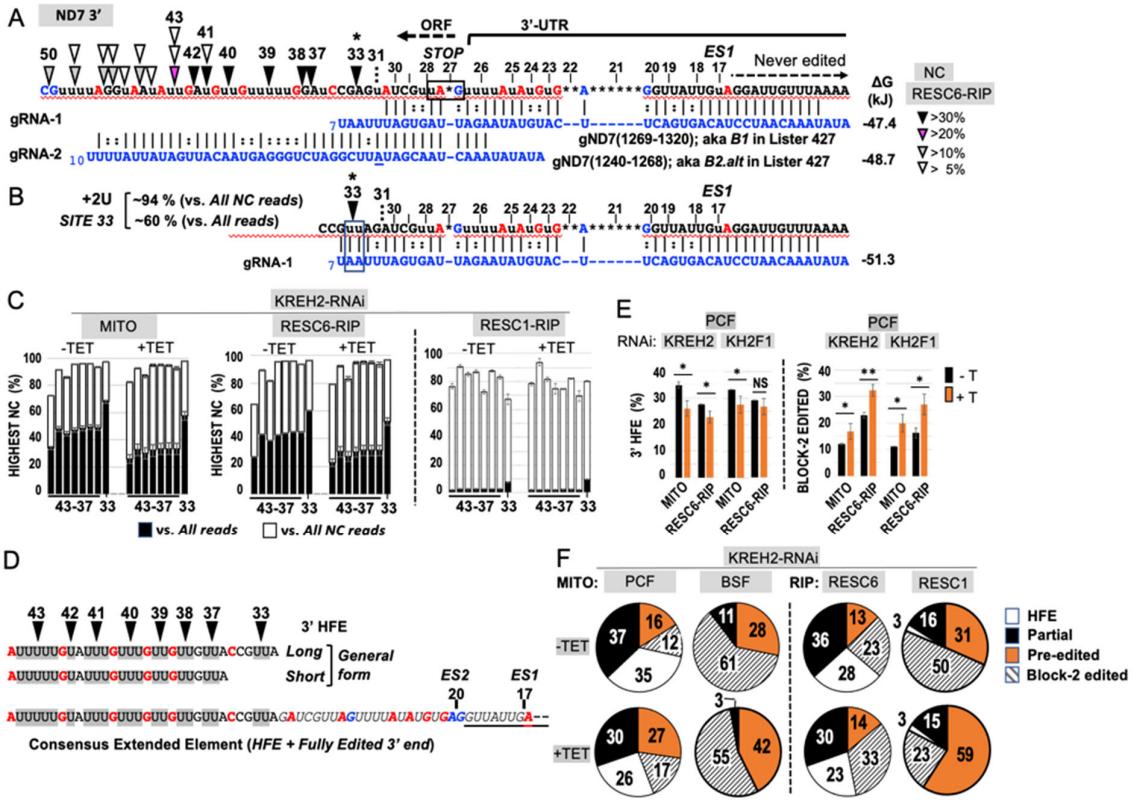
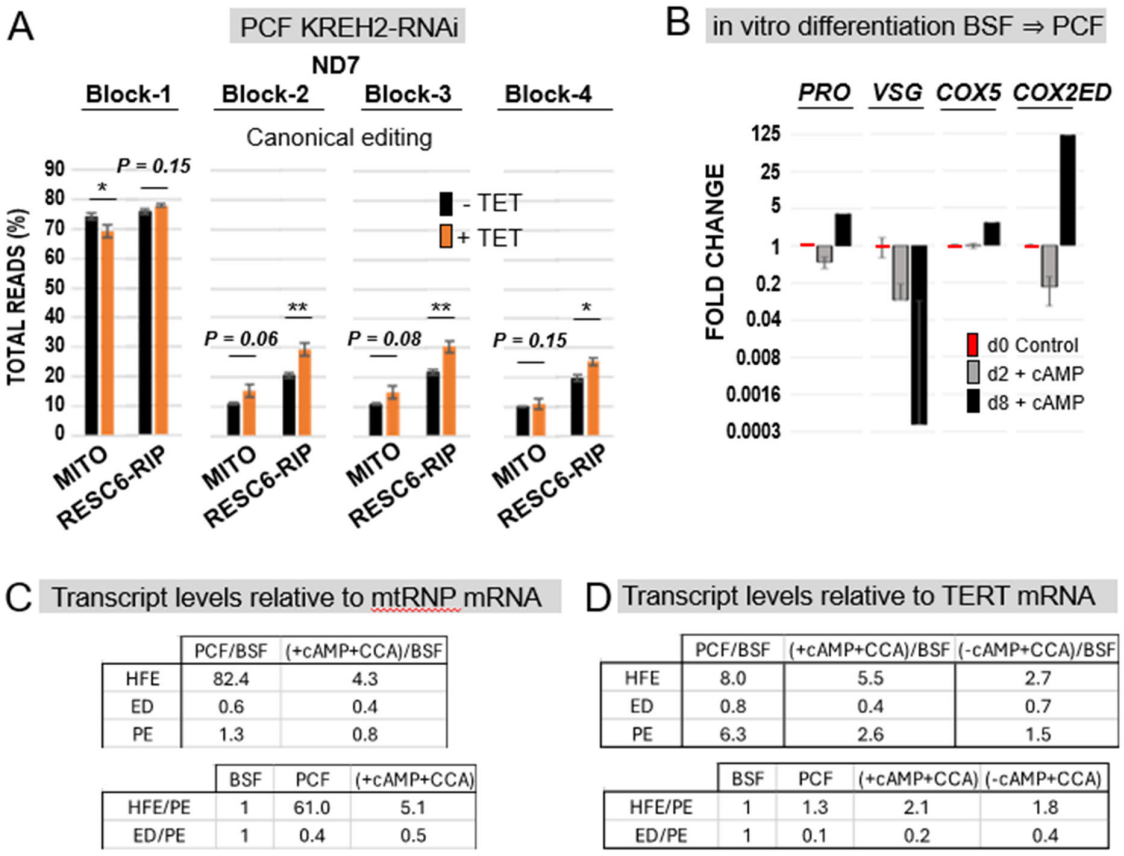

A
ND7 3'
ORF
3'-UTR
STOP
ES1
Never edited
ΔG (kJ)
NC
RESC6-RIP
gRNA-1
gRNA-2
gND7(1269-1320); aka B1 in Lister 427
gND7(1240-1268); aka B2.alt in Lister 427
-47.4
-48.7
>30%
>20%
>10%
> 5%
B
+2U
SITE 33
~94 % (vs. All NC reads)
~60 % (vs. All reads)
-51.3
C
KREH2-RNAi
MITO
RESC6-RIP
RESC1-RIP
-TET
+TET
HIGHEST NC (%)
vs. All reads
vs. All NC reads
D
3' HFE
Long
Short
General form
ES2
ES1
Consensus Extended Element (HFE + Fully Edited 3' end)
E
PCF
RNAi: KREH2 KH2F1
3' HFE (%)
BLOCK-2 EDITED (%)
- T
+ T
NS
F
KREH2-RNAi
MITO: PCF BSF
RIP: RESC6 RESC1
HFE
Partial
Pre-edited
Block-2 edited

Figure 5. An abundant non-canonical 3′ high-frequency element (HFE) sequence in the ND7 3′ editing domain in PCF and BSF, and its modulation by REH2C proteins. (A) Canonically edited ND7 3′ terminus. Color-coded letters are just 3′ of sites for sites requiring insertion (red), deletion (blue), or no changes (black). ORF, 3′ UTR, and never-edited regions are indicated. The first editing site (*ES1*) is at position 17, counting from the 3′ end. Illumina sequenced gRNA isoforms: gRNA-1 *B1* and gRNA-2 *B2.alt* in strain Lister 427 (31) correspond to gRNA-1 gND7(1269-1320) and gRNA-2 gND7(1240-1268**)** in strain EATRO 164 (22), with identical guiding function at block 1 and block 2, respectively. These guides produced the best match with the ND7 3′ edited pattern in this study and prior Sanger sequencing in our strain (31). Color-coded arrowheads indicate sites with a dominant NC read representing >30% (black) or less in the color-coded scale of all reads in RESC6-RIPs in uninduced (-Tet) cells containing a KREH2-RNAi construct. **(B)** Canonical and non-canonical guiding potential of the initiator gRNA-1 *B1*. Alternative uninterrupted alignment of the 3′ terminus in gRNA-1 would involve non-canonical addition of +2 U at site 33 and 0 U at site 31. **(C)** Actual percentage of the dominant NC read at each indicated site versus all reads (black) or versus all NC reads (white) in mtRNA and RESC6 or RESC1-RIPs in PCF. **(D)** Consensus 3′ High-Frequency Element (3′ HFE) made by the dominant NC reads at sites 33-43. The top two 3′ HFE isoforms found in all samples examined show the dominant NC reads in gray. General 3′ HFE long or short forms, where "n" represents any T nucleotide number at sites 34-36. Bottom: ~62-nt extended 3' element, including the 3′ HFE and 3′ terminal fully edited sequence. **(E)** Frequency of 3′ HFE (*left*) or canonically edited block 2 (*right*) in mtRNA or RESC6-RIPs in PCF ± KREH2-RNAi or KH2F1-RNAi. **(F)** Frequency of 3′ HFE, and either block 2 read type: canonically edited, pre-edited, and other NC ("partial") in PCF and BSF ± KREH2-RNAi cells. ±Tet (also labeled ±T). Statistical analyses are as in **Fig. 2**.



	PCF/BSF	(+cAMP+CCA)/BSF
HFE	82.4	4.3
ED	0.6	0.4
PE	1.3	0.8

	BSF	PCF	(+cAMP+CCA)
HFE/PE	1	61.0	5.1
ED/PE	1	0.4	0.5

	PCF/BSF	(+cAMP+CCA)/BSF	(-cAMP+CCA)/BSF
HFE	8.0	5.5	2.7
ED	0.8	0.4	0.7
PE	6.3	2.6	1.5

	BSF	PCF	(+cAMP+CCA)	(-cAMP+CCA)
HFE/PE	1	1.3	2.1	1.8
ED/PE	1	0.1	0.2	0.4

Figure 6. Relative changes in ND7 3′ edited blocks 1-4 upon KREH2-RNAi knockdown and in 3′ HFE upon *in vitro* differentiation. (A) RNA-seq analyses of the total reads (%) of canonically edited ND7 3′ domain blocks 1 through 4 in mtRNA or RESC6-RIPs in PCF ± KREH2-RNAi ±Tet. **(B)** qPCR analyses of relative levels for hallmark transcripts during *in vitro* differentiation. BSF strain Lister 427 cells before (day 0; d0) and after (day 2, d2, or day 8, d8) *in vitro* differentiation to PCF via 2 days of incubation at 37°C in BSF media +pCPTcAMP (cAMP) and 6 more days of incubation in PCF media with citrate/cis-aconitate at 26°C (full details of in vitro differentiation protocol in Materials and Methods). **(C-D)** Independent *in vitro* differentiation experiments and qPCR analyses (from two different labs; see Materials and Methods for more details) of 3′ HFE, canonically edited, and pre-edited ND7 normalized to mt-RNA polymerase mRNA (C) or TERT (D) as reference transcripts. Ratios in parent cells, PCF/BSF, or post/prior differentiation (±cAMP +CCA)/BSF are compared for the independent experiments in Lister 427 cells grown separately in two different labs for decades.

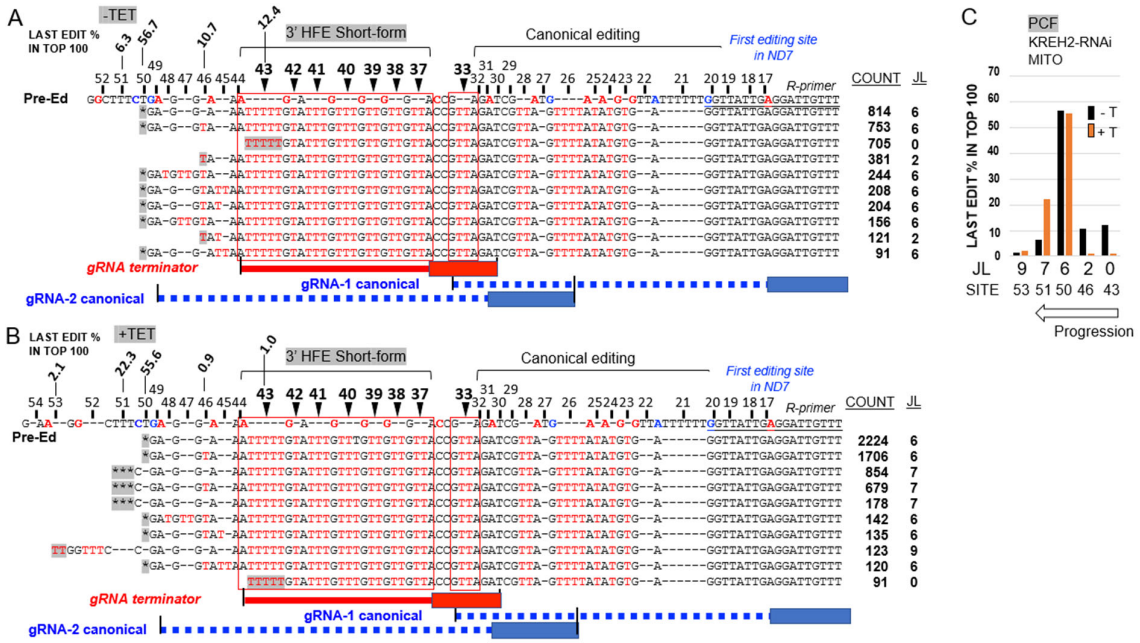
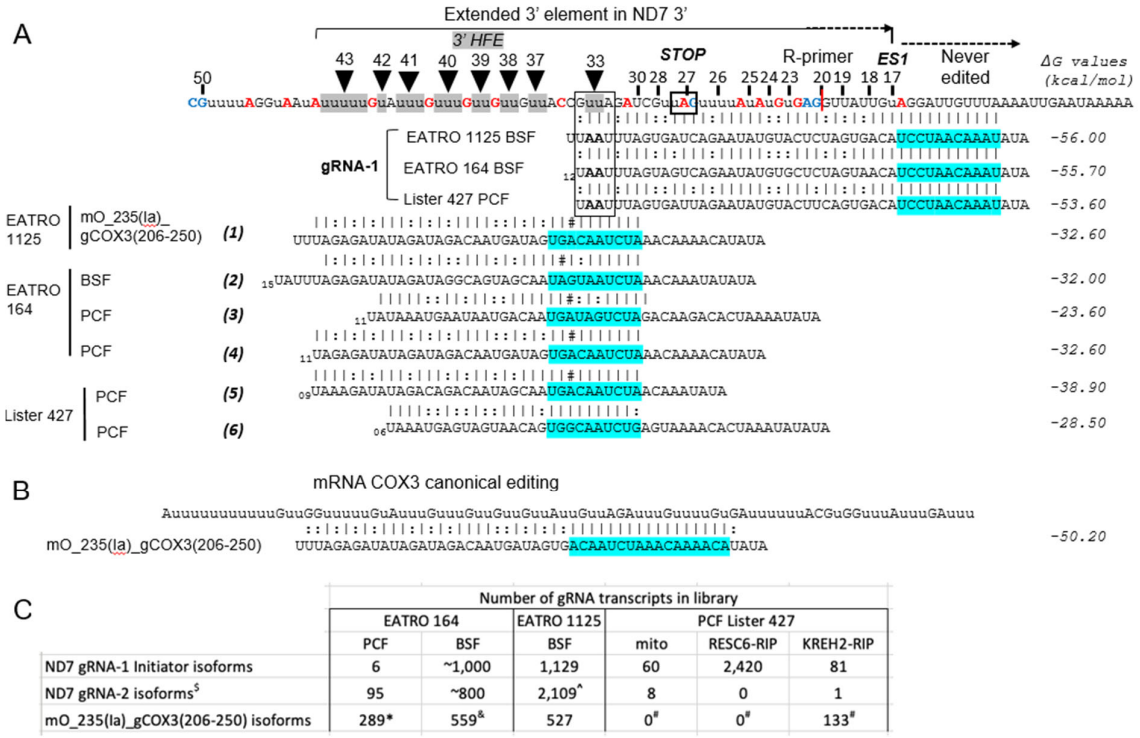


Figure 7. Most frequent ND7 amplicons that contain the non-canonical 3′ HFE sequence. Sequence alignment of the top ten amplicons with the short-form 3′ HFE (sites 37-43) in a representative sample of RESC6-IP in PCF KREH2-RNAi **(A)** uninduced -Tet or **(B)** induced +Tet. The last edit (gray) in each unique sequence is indicated as a percentage in the top 100 amplicons. Sequence 5′ to the last edit is pre-edited or includes a variable length non-canonical editing junction. The count of each amplicon in the top 100 amplicons in that sample examined is shown. The cognate initiator gRNA-1 and gRNA-2 with an anchor (box) and guiding region (dashed line) in blue, a terminator gRNA matching the 3′ HFE (straight line) in red, and the first canonical editing site in ND7 3′ (position 17) are depicted. Color-coded letters are as in prior figures. **(C)** Junction length (JL) from panels A and B, counted as the number of sites containing any edits upstream of site 43. The site containing the last edit in the junction is indicated. Amplicons carry canonically edited sequence 3′ to the 3′ HFE. Junction length (JL) analyses of other samples were also performed (**Supplementary Figs. S9-S11**).



C

	Number of gRNA transcripts in library					
	EATRO 164		EATRO 1125	PCF Lister 427		
	PCF	BSF	BSF	mito	RESC6-RIP	KREH2-RIP
ND7 gRNA-1 Initiator isoforms	6	~1,000	1,129	60	2,420	81
ND7 gRNA-2 isoforms$	95	~800	2,109^	8	0	1
mO_235(Ia)_gCOX3(206-250) isoforms	289*	559&	527	0#	0#	133#

Figure 8. Coordinated non-canonical usage of the initiator gRNA-1 and a novel terminator guide may install the extended 3′ element in ND7 3′. **(A)** The extended 3′ element from sites 17 to 43 in ND7 3′, as in **Fig. 5D,** aligned with gRNA-1 isoforms in strains BSF EATRO 1125, PCF EATRO 164, and PCF Lister 427 (used in this study) with non-canonical +2U at site 33 (boxed). A novel gRNA mO_235(Ia)_gCOX3(206-250) identified in strain EATRO 1125 (transcript #1) or isoforms in other strains (transcripts #2-6) match the 3′ HFE except for a single C:A mismatch (either C in a CC doublet) at site 35 or 36 in mRNA. Guide isoform #6 in strain Lister 427 has no matches but ends prematurely at site 41 in the 3′ HFE. **(B)** Sequence alignment of gRNA mO_235(Ia)_gCOX3(206-250) with its cognate mRNA COX3. Predicted ΔG values in kcal/mol) of the duplexes are indicated. The original stop codon created by gRNA-1 in ND7 3′, a never-edited sequence, the reverse (R) primer 5′ end (red line), and anchor regions (blue) are annotated. **(C)** Number of gRNA transcripts in available databases in PCF and BSF. Initiator gRNA-1 isoforms in EATRO 164, gA6 (774-822), and Lister 427, gA6 *B1.alt*, have the same guiding capacity as in EATRO 1125. gRNAs in EATRO 164 and EATRO 1125 were determined in total mtRNA; gRNAs in Lister 427 were determined in RESC6-IPs and available alignments online (31).

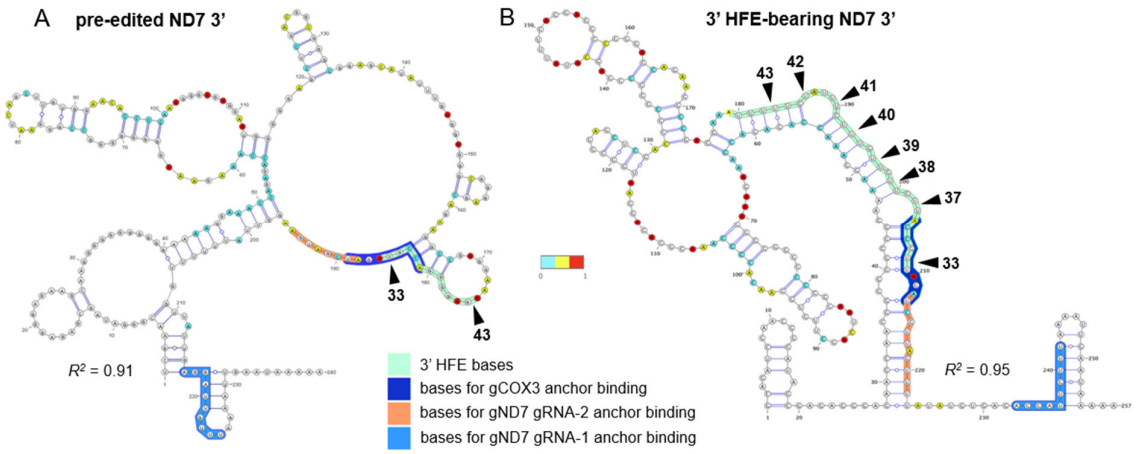


Figure 9. Structural models determined by DMS-MaPseq of the (A) pre-edited (PE) and (B) 3′ HFE-containing ND7 3′ terminus. Nucleotides are color-coded by normalized DMS signal. The 3′ HFE is highlighted in a green box; nucleotides used in binding the anchor region in cognate ND7 gRNA-1 and gRNA-2 and non-cognate gCOX3 are highlighted in blue, orange, and royal blue, respectively. Both structures carry a complete anchor binding site for cognate gND7 gRNA-1. Only the 3′ HFE-containing structure carries a complete element (sites 33-43) and complete anchor binding sites for cognate gND7 gRNA-2 and non-cognate gCOX3. R^2 is Pearson's R^2 in structures from n = 2 experiments. DMS reactivity is calculated as the ratiometric DMS signal per position normalized to the highest number of reads in the displayed region, which is set to 1.0.

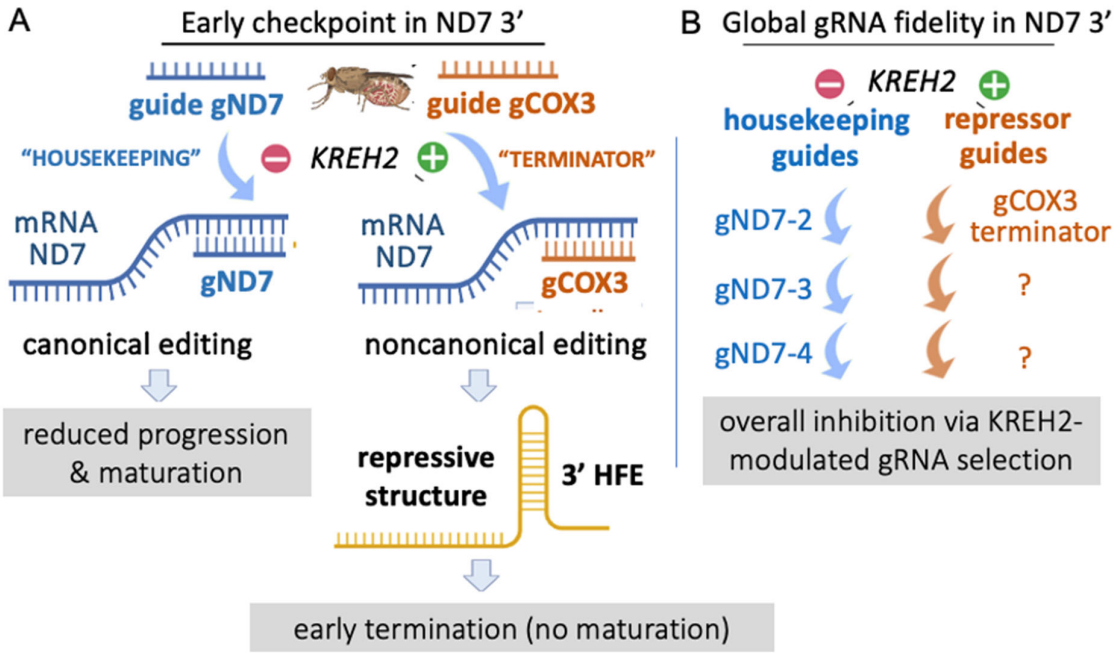


Figure 10. Model of REH2C-dependent developmental ND7 editing regulation via modulation of gRNA selection and RNA structure. In this model, KREH2 maintains a state of ND7 3′ editing repression in PCF involving modulation of gRNA selection and installation of a proposed RNA structure. PCF-specific repression was defined at two levels: **(A)** at a major early editing checkpoint where KREH2 exhibits concurrent positive (+) control on usage of a novel "terminator" guide (gCOX3) and negative (-) control on usage of the housekeeping cognate gRNA-2 (gND7). KREH2 promotes preferential usage of the terminator over gND7. Editing using the terminator installs a non-canonical 3′ HFE in ~30% of the ND7 transcriptome, which has two main effects: first, it derails all upstream canonical editing, and second, it forms a stable fold, potentially hindering downstream gRNA-2 binding and "repair" via re-editing to replace the 3′ HFE with canonical sequence, and **(B)** globally, with gRNA fidelity changes by KREH2 promoting "repressive" non-cognate guiding and inhibiting housekeeping cognate guiding at most ND7 3′ sites examined. Thus, inhibition of ND7 3′ editing maturation in PCF is mainly at the early checkpoint but also occurs globally in transcripts that did not receive the 3′ HFE. KREH2-mediated repression in ND7 3′ was not observed in BSF cells.