Figures and figure supplements

Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking

Pornchai Kaewsapsak et al
Figure 1. APEX-RIP in mitochondria. (A) Overview of the APEX-RIP workflow. Live cells expressing APEX2 (grey ‘pacmen’) targeted to the compartment of interest (here, the mitochondrial matrix) are incubated with the APEX substrate biotin-phenol (BP; red B: biotin). A one-minute pulse of H$_2$O$_2$ initiates biotinylation of proximal endogenous proteins (Rhee et al., 2013), which are subsequently crosslinked to nearby RNAs by 0.1% formaldehyde. Following cell lysis, biotinylated species are enriched by streptavidin pulldown, and coeluting RNAs are analyzed by qRT-PCR or RNA-Seq. IMM: inner mitochondrial membrane. (B) Imaging APEX2 biotinylation in situ. HEK 293T cells expressing V5-tagged mito-APEX2 were biotinylated using the APEX-RIP workflow, fixed, and stained as indicated. The bottom row is a negative control in which H$_2$O$_2$ treatment was omitted. Scale bars, 10 μm. TOM20 is Figure 1 continued on next page.
a mitochondrial outer membrane protein; neutravidin staining detecting biotinylation. (C) In situ biotinylation of the mitochondrial matrix proteome requires mito-APEX2, BP, and H2O2. Streptavidin blot analysis of whole cell lysates prepared following the protocol described in (A), or after omitting components of the APEX reaction. Arrowheads denote endogenous biotinylated proteins (Chapman-Smith and Cronan, 1999). Anti-V5 blot (bottom) detects expression of mito-APEX2. (D–E) mito-APEX-RIP efficiently recovers the mitochondrial transcriptome. (D) Gene-level RNA-Seq analysis of mito-APEX-RIP; data are the average values of three experimental replicates. Fold change is defined as (FPKMpost-enrichment/FPKMpre-enrichment); dashed lines indicate significance thresholds for fold enrichment (determined by ROC analysis, see Materials and methods) and p-values calculated by CuffDiff2 (Trapnell et al., 2013). Mitochondrial genomes encode 13 mRNAs, two rRNAs and 22 tRNAs (red). Note that three mitochondrial tRNA genes, MT-TH, MT-TL2, and MT-TG, were also enriched. See Supplementary file 1A. (E) Nucleotide-level RNA-Seq analysis of mito-APEX-RIP, mapped to the human mitochondrial genome (innermost circle). Outermost circle: reads from the full APEX-RIP protocol; middle circle: reads from the negative control. Note the enrichment of several mitochondrially-encoded tRNAs and the D-loop leader transcript. Ribosomal RNAs were removed during library preparation (see Materials and methods). See also: Figure 1—figure supplements 1,2.
Figure 1—figure supplement 1. Optimization of APEX-RIP protocol. (A) Top: Alternative labeling and crosslinking protocols. In protocol I, cells are crosslinked with formaldehyde (FA) and quenched with Glycine (Gly) prior to the introduction of biotin-phenol (BP) and the initiation of APEX-catalyzed biotinylation with H$_2$O$_2$. In Protocol II., live cells are incubated in BP, and in situ biotinylation is initiated prior to FA crosslinking. In both cases, cell lysis, streptavidin enrichment and RNA purification proceed as described (see Materials and methods). Bottom: qRT-PCR analysis comparing Protocols I and II. Negative control experiments replace mito-APEX with mito-GFP, or omit BP and H$_2$O$_2$. All constructs were APEX1 derivatives, transiently expressed in HEK 293T cells. Data are the means of three replicates ± one standard deviation. (B) RNA-Seq analysis of RNAs enriched by protocol I. Although all 13 mitochondrially-encoded mRNAs (green) were enriched, these were accompanied by substantial contaminating RNAs, including XIST and MAN2C1. Data from one experimental replicate are shown. (C) qRT-PCR analysis comparing Protocols I and II., including off-target controls designed using the
results from RNA-Seq in (B). Note the superior enrichment obtained using Protocol II. Cells in this experiment stably-expressed mito-APEX2. This protocol and cell line were used to collect all data in Figure 1. Data are the means of four replicates ± one standard deviation. Significance testing: Student's two-tailed t-test.

DOI: https://doi.org/10.7554/eLife.29224.003
Figure 1—figure supplement 2. Reproducibility of the mito-APEX2 RIP experiment. Pairwise correlations between global RNA abundances (FPKM) in biological replicates (A) prior to enrichment, and (B) following mito-APEX2-RIP.

DOI: https://doi.org/10.7554/eLife.29224.004
Figure 2. APEX-RIP mapping of the nuclear-cytoplasmic RNA distribution. (A) Fluorescence imaging of nuclear and cytosol-targeted APEX fusion constructs. HEK 293T cells expressing the indicated constructs (‘NLS,’ nuclear localization signal; ‘NES,’ nuclear export signal) were labeled with biotin-

Figure 2 continued on next page
Figure 2 continued

phenol, crosslinked and stained as indicated. Scale bars, 10 μm. DAPI is nuclear stain. (B) APEX-RIP recovers known nuclear and cytosolic standard RNAs, defined here as long noncoding RNAs (nuclear markers, blue) and RNAs proximal to the Endoplasmic Reticulum (cytoplasmic markers, red—defined by (Jan et al., 2014), with measured p-values ≤ 0.05—see Materials and methods). Top: APEX2–NLS-RIP enriches nuclear standards. Middle: APEX2–NES-RIP enriches cytoplasmic standards. Bottom: Combined analysis of the APEX2–NLS and APEX2–NES RIP experiments distinguish the two classes. Fold changes are defined as (FPKM_{post-enrichment}/FPKM_{pre-enrichment}), combined fold change as (FPKM_{NLS-post-enrichment}/FPKM_{NES-post-enrichment}). Dotted line indicates the significance threshold for nuclear localization. (C) Global analysis of nuclear and cytoplasmic RNA localization by combined APEX2–NLS and APEX2–NES RIP. Vertical dashed lines indicate the cutoffs for nuclear and cytosolic RNAs. Horizontal dash line indicates p-value=0.05. Top histogram illustrates the distribution of RNAs with p-value=5 × 10^{-5}, which are boxed in gray in the scatter plot. The average of data from three biological replicates are shown. See Supplementary file 2A-B. (D) APEX-RIP reveals classes of RNAs with canonical and noncanonical nuclear-cytoplasmic distributions. Left: the same data as in (C), separately parsed into mRNAs (top) and lncRNAs (bottom). Right: read density plots of example RNAs from each class that exhibit stereotypical and atypical localization. EEF1A1 and C1orf63 are mRNAs; XIST and SNHG5 are lncRNAs. For each gene, a common y-scale is used for all read tracks. SnoRNAs encoded in the SNHG5 gene body are indicated as gray rectangles. (E) Venn diagrams comparing APEX-RIP and fractionation-based RNA datasets (Sultan et al., 2014). (F) Nuclear APEX-RIP is more sensitive than is biochemical fractionation. Left: Specificity of the APEX-RIP and nuclear RNA datasets (Sultan et al., 2014). Off-target RNAs were defined as actively translated ER-proximal mRNAs (Jan et al., 2014). Right: Recall of nuclear standard RNAs, defined as a set of 827 lncRNAs annotated by GENCODE hg19 with average pre-enrichment FPKM ≥ 1.0. See also: Figure 2—figure supplements 1–2.

DOI: https://doi.org/10.7554/eLife.29224.005
**Figure 2—figure supplement 1.** Characterization of APEX2 fusion constructs. HEK 293 T cells stably expressing the indicated constructs (right) were labeled and crosslinked via Protocol II (Figure 1—figure supplement 1A). Cell lysates were analyzed by SDS-PAGE, blotting with streptavidin-HRP, anti-V5 and anti-FLAG. L: ladder; U: untransfected HEK 293T cells. Anti-V5 and anti-FLAG blots (bottom left) measure fusion construct expression.

DOI: https://doi.org/10.7554/eLife.29224.006
A
Correlation between pre-enrichment replicates of APEX2-NLS RIP experiment

B
Correlation between post-enrichment replicates of APEX2-NLS RIP experiment

C
Correlation between pre-enrichment replicates of APEX2-NES RIP experiment

D
Correlation between post-enrichment replicates of APEX2-NES RIP experiment
Figure 2—figure supplement 2. Reproducibility of nuclear–cytoplasmic APEX-RIP experiments. Pairwise correlations between global RNA abundances (FPKM) in biological replicates of the APEX2–NLS-RIP experiment, (A) prior to, and (B), following enrichment, and similarly, of the APEX2–NES-RIP experiment, (C) prior to, and (D) following enrichment.

DOI: https://doi.org/10.7554/eLife.29224.007
Determination of cutoffs for APEX-RIP data by ROC analysis (see Materials and methods). (C) APEX protein localization isn’t perturbed during the APEX-RIP labeling and crosslinking protocol. APEX2-NLS HEK293T cells expressing a whole-cell GFP marker were processed using the optimized protocol.

**Figure 2—figure supplement 3 continued on next page**

---

**Kaewsapsak et al. eLife 2017;6:e29224. DOI: https://doi.org/10.7554/eLife.29224**

12 of 22
APEX-RIP protocol (bar diagram). Cells were fixed prior to, and following the mild formaldehyde crosslinking step (arrowheads), and imaged by immunofluorescence. Left: Representative images for cells fixed at each time point. Right: average nuclear-vs-cytosolic biotinylation signal ratio measured at both time points; each dot represents an individual cell (n = 30). p-value: student’s paired t-test. (D—E) APEX-RIP recapitulates the results of nuclear-cytosolic fractionation-sequencing. (D) Global analysis of nuclear and cytoplasmic RNA localization [FPKM$_{nuclear}$/FPKM$_{cytosolic}$] by biochemical fractionation-sequencing of HEK293 cells (Sultan et al., 2014). Horizontal dash line indicates p-value = 0.05. Top histogram illustrates the distribution of RNAs with p-value $\approx$ $5 \times 10^{-5}$, which are boxed in gray in the scatter plot. Average data from two biological replicates are shown. (E) Distribution of RNAs based on enrichment of nuclear over cytosolic fractionation, similar to the bottom panel of Figure 2B (see Materials and methods). (F–G) Determination of cutoffs for fractionation data (Sultan et al., 2014) by ROC analysis (see Materials and methods).

DOI: https://doi.org/10.7554/eLife.29224.008
Figure 3. APEX-RIP at the Endoplasmic Reticulum membrane. (A—B) Schematics summarizing alternating ER-targeting strategies. (A) HRP, targeted to the ER lumen with a KDEL sequence, biotinylates proteins within the ER. Red B: biotin. Red X’s: chemical crosslinks induced by 0.1% formaldehyde treatment. Note that RNAs enriched by this approach may reside at the cytosolic face of the ER, or at the nuclear lamina, as shown. (B) APEX2, targeted to the ER lumen, facilitates RNA enrichment at the cytosolic face of the ER. (C) Representative images show the enrichment of RNAs with either (HRP+H₂O₂) or (ERM+H₂O₂) treatment. (D) Western blot analysis confirms the enrichment of ER-associated RNAs with H₂O₂ treatment. (E) Quantification of RNAs enriched with either (HRP+H₂O₂) or (ERM+H₂O₂) treatment. The statistical significance was determined using the Student’s t-test (two-tailed) with a p-value of less than 0.05. The data are presented as mean ± SD. Figure 3 continued on next page.
Figure 3 continued

to the ER membrane (ERM) by fusing it to the transmembrane segment of rabbit P450 C1, biotinylates proteins proximal to the cytosolic face of the ER. (C) Imaging of biotinylation from HR-P-KDEL and ERM-APEX2 catalyzed reactions. HEK293T cells stably expressing HR-P-KDEL or ERM-APEX2 were labeled with BP, fixed and imaged as in Figure 1B. Scale bars, 10 μm. Anti-RCN2 was used to mark ER lumen. (D) Streptavidin blot detection of resident ER proteins biotinylated by HR-P-KDEL, as in Figure 1C. Arrowheads denote endogenously biotinylated proteins (Chapman-Smith and Cronan, 1999). (E) qRT-PCR analysis, comparing the specificities of the labeling schemes shown in (A–B). Target and off-target genes were selected using previously-reported RNA abundances at the ER membrane (Jan et al., 2014). Cytosolic APEX2 (APEX2–NES, as in Figure 2) serves as a negative control. Data are the mean of four replicates ± one standard deviation. Significance testing: Student’s two-tailed t-test.
DOI: https://doi.org/10.7554/eLife.29224.009
Figure 3—figure supplement 1. Further optimization of the APEX-RIP protocol; additional HRP-KDEL RIP data. (A) Addition of a radical quenching step between APEX2 labeling and formaldehyde crosslinking improves APEX-RIP specificity. Top: schematic of the revised labeling–crosslinking workflow. The radical quenchers used were Trolox and ascorbic acid; these were included in the 0.1% FA and Gly steps regardless of whether the discrete radical quenching step is included. Bottom: qRT-PCR analysis of the ER-APEX-RIP experiment with and without this additional quenching step, as in (Figure 1—figure supplement 1C). Data are the mean of four replicates, ± one standard deviation. Significance testing: Student's two-tailed t-test. (B–C) Correlation of RNA abundance (in FPKM) for each replicate. (B) and (C) are pre- and post-enrichment replicates, respectively. Given the low correlation between post-enrichment replicate three and other replicates, only data from replicates 1 and 2 were included in further analysis. (D) Histograms showing the distribution of RNA enrichment values—log2 fold change (defined as FPKM_{post-enrichment}/FPKM_{pre-enrichment}), for all RNAs with p-values ≤ 0.05. Green denotes mRNAs encoding known secretory proteins and red denotes mRNAs encoding non-secretory proteins. Significance testing...
cutoffs were determined using ROC analysis (see below). Known secretory and non-secretory standard mRNAs were cataloged as described (see Materials and methods). (E–F) Determination of significance thresholds ratio cutoffs by ROC analysis (see Materials and methods). DOI: https://doi.org/10.7554/eLife.29224.010
Figure 3—figure supplement 2. Additional analysis comparing HRP-KDEL RIP data and other ER-RNA data sets. (A) Gene ontology (GO) analysis of mRNAs that were enriched in ER datasets, but which lack secretory annotation. The number of RNA species in each class is indicated to the right of the corresponding bar.

Left: non-secretory mRNAs enriched by biochemical fractionation (99 mRNAs, 9% of total enriched mRNAs) (Reid and Nicchitta, 2012) predominantly exhibit nuclear and mitochondrial annotation. Right: non-secretory mRNAs enriched by HRP-KDEL RIP (87 mRNAs, 3.5% of total APEX-RIP-enriched mRNAs) have predominantly cytosolic annotation, with a minority exhibiting nuclear annotation. All terms shown have $p<0.05$, as assessed using DAVID (Huang et al., 2009).

(B) APEX-RIP recovers low-abundance targets with greater efficiency than do alternative approaches. Left: RNAs enriched by ER-APEX-RIP exhibit a broader range in starting abundances than do RNAs recovered by ER biochemical fractionation or ER-proximity restricted ribosome profiling. Right: Table summarizing the lower bound (i.e. the lowest starting FPKM) for 95% of the RNA species enriched by each method.

DOI: https://doi.org/10.7554/eLife.29224.011
Figure 4. Mapping the ER-proximal transcriptome with APEX–RIP. (A) Global analysis of RNA localization at the Endoplasmic Reticulum. Fold change = (FPKM_{post-enrichment}/FPKM_{pre-enrichment}). Horizontal dashed line indicates p-value = 0.05. Top histogram illustrates the distribution of RNAs with p-
value = $5 \times 10^{-5}$, which are boxed in gray in the scatter plot. Average data from two biological replicates are shown. Standard mRNAs encoding known secretory and non-secretory proteins are highlighted in red and blue, respectively (see Materials and methods). (B) Classification of APEX-RIP enriched, ER-associated RNAs. Collectively, all classes of non-coding RNAs constitute 6.7% of enriched genes (178 of 2672 RNAs). (C) Specificity analysis for protein-coding mRNAs in the APEX-RIP-derived ER-associated RNA list. The total number of RNA species in each condition is indicated above each column. 96.5% of the 2494 APEX-RIP ER-enriched mRNAs exhibit some form of secretory annotation (as predicted by Phobius, TMHMM, SignalP, or GOCC, see methods), whereas only 53.8% of total mRNAs expressed in HEK293T cells (FPKM $\geq 1.0$) are similarly classified (left). (D) Target recall of ER APEX-RIP exceeds those of proximity-restricted ribosome profiling (Jan et al., 2014; see Supplementary file 3D) and biochemical fractionation (Reid and Nicchitta, 2012; see Supplementary file 3C). See also: Supplementary file 3E. (E) Venn diagram comparing RNA datasets. Note that all enriched RNAs in Reid et al. ER fractionation-Seq dataset were mRNAs. See also: Figure 3—figure supplements 1–2.

DOI: https://doi.org/10.7554/eLife.29224.012
Figure 5. APEX-RIP reveals RNAs with potentially novel localization. (A) Many mitochondrial transmembrane proteins appear to be translated at the ER. mRNAs encoding mitochondrial proteins (defined by GOCC and MitoCarta 1.0 (Ashburner et al., 2000; Pagliarini et al., 2008)) with predicted localization were captured in both ER and KDEL-RIP conditions. (B) The fraction of mRNA population that is enriched by APEX2-NES is shown for GOCC-annotated mRNAs encoding mito-proteins, ER-associated materials, and ER-mitochondrial (ER-M) proteins. (C) Schematic representation of nuclear proteins, nuclear lamination, nuclear lamina, nuclear RNAs, lamina-associated RNAs, ER proteins, ER-associated RNAs, cytosol, and nucleus. (D) Summary of APEX-RIP nuclear RNAs, 5299 RNAs: 2776 mRNAs, 607 lincRNAs, 1961 other ncRNAs. 2572 APEX-RIP ER-associated RNAs: 2494 mRNAs, 28 lincRNAs, 150 other ncRNAs. 2231 RNAs: 2109 mRNAs, 17 lincRNAs, 105 other ncRNAs. 441 RNAs: 385 mRNAs, 45 other ncRNAs. P-value (all RNAs) = 2.74 x 10^-7, P-value (lincRNAs) = 1.06 x 10^-4. 

[Figure 5 continued on next page]
transmembrane helices (predicted by TMHMM (Krogh et al., 2001); red distribution) are preferentially enriched by HRP-KDEL APEX-RIP, relative to those encoding mitochondrial proteins lacking transmembrane domains (gray distribution). See Supplementary file 3A. (B) Mitochondrial proteins encoded by ER-proximal mRNAs are enriched for outer mitochondrial membrane (OMM) destined proteins, and de-enriched for matrix and intermembrane space (IMS) destined proteins. Predicted sub-mitochondrial localization of all GOCC-annotated mitochondrial proteins (left), those with mRNAs enriched by APEX2–NES (middle), and those enriched by HRP–KDEL (right). IMM: Inner mitochondrial membrane. The total number of mRNA species with annotated mitochondrial sublocalization is indicated above each column. See Supplementary file 4A. (C) Scheme for identifying putative RNAs associated with the nuclear lamina. Since subsets of laminar proteins should be biotinylated both by APEX2-NLS (top right) and by HRP-KDEL (bottom right and Figure 3A), we can intersect these two datasets to obtain a candidate list of nuclear lamina-localized RNAs. Notation as in Figure 3A. (D) Venn diagram identifying putative lamina-associated RNAs, defined as the overlap between HRP-KDEL- and APEX2-NLS-enriched RNAs. See also: Supplementary file 4B. Significance testing: hypergeometric test.

DOI: https://doi.org/10.7554/eLife.29224.013