ARTICLES
https://doi.org/10.1038/s41589-019-0368-5
nature chemical biology

Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling

Pengchong Wang[1,2,3,4,5,12], Wei Tang[6,7,12], Zeyao Li[1,8,12], Zhongyu Zou[2,3,4,5], Ying Zhou[2,3,4,5], Ran Li[6,7], Tianyu Xiong[9], Jianbin Wang[1,7,10*] and Peng Zou[2,3,4,5,7,11*]

RNA molecules are highly compartmentalized in eukaryotic cells, with their localizations intimately linked to their functions. Despite the importance of RNA targeting, our current knowledge of the spatial organization of the transcriptome has been limited by a lack of analytical tools. In this study, we develop a chemical biology approach to label RNAs in live cells with high spatial specificity. Our method, called CAP-seq, capitalizes on light-activated, proximity-dependent photo-oxidation of RNA nucleobases, which could be subsequently enriched via affinity purification and identified by high-throughput sequencing. Using this technique, we investigate the local transcriptomes that are proximal to various subcellular compartments, including the endoplasmic reticulum and mitochondria. We discover that messenger RNAs encoding for ribosomal proteins and oxidative phosphorylation pathway proteins are highly enriched at the outer mitochondrial membrane. Due to its specificity and ease of use, CAP-seq is a generally applicable technique to investigate the spatial transcriptome in many biological systems.

The subcellular targeting of RNA molecules is crucial for their biological functions. This is most evident in eukaryotic cells where RNA can be segregated into specific cytoplasmic compartments to enable local protein synthesis[1,2], which is fundamental to a wide range of biological processes, including cell proliferation[3], embryonic development[4] and neuronal plasticity[5,6]. In addition to functioning as coding molecules, localized RNA can also play regulatory or architectural roles. For example, nuclear noncoding RNAs can interact with chromatin to regulate gene transcription, maintain chromatin conformation or participate in the formation of membrane-less organelles such as nuclear speckles[7].

Our knowledge of RNA targeting has greatly advanced in the past, due to the development of analytical techniques to isolate and image the spatial transcriptome. Subcellular fractionation coupled with microarray analysis or high-throughput sequencing has produced valuable information about the local transcriptome at the cellular membrane[8], endoplasmic reticulum (ER)[9], mitochondria[10,11] and neuronal processes[12,13]. Alternatively, cellular RNA molecules could be directly visualized under the microscope, by RNA reporter assay[14], fluorescence in situ hybridization (FISH)[15,16] and in situ sequencing[17,18]. Modern developments in instrumentation have expanded the throughput of imaging analysis from a handful of RNAs to over a thousand transcripts, revealing distinct patterns of RNA expression and distribution in human fibroblast cells (MERFISH[16] and FISSEQ[17]) and neurons (STARmap[18]). Recently, peroxidase-mediated protein biotinylation (APEX[19]) has been combined with protein–RNA crosslinking methods (APEX-RIP[20] and Proximity-CLIP[21]) to investigate the subcellular transcriptome of the nucleus, the mitochondrial matrix, the ER membrane and the cell–cell interface. Each of the above methods has its own merits in spatial resolution (for example, imaging and genetic targeting), analysis throughput (for example, RNA sequencing) and compatibility with live cells (for example, RNA reporter assay).

In this study, we present a new approach that combines the high spatial specificity of genetic targeting with the high throughput of sequencing analysis. Our method, called CAP-seq—chromophore-assisted proximity labeling and sequencing—capitalizes on a genetically encoded photosensitizer that mediates the proximity-dependent photo-oxidative conjugation of an amine probe to RNA nucleobases in live cells. These tagged RNA molecules are subsequently purified and sequenced. We applied CAP-seq technique to profile RNAs at several subcellular compartments. In the mitochondrial matrix, CAP-seq achieved exceptional spatial resolution and sensitivity, capturing all of its 13 messenger RNAs while excluding highly abundant cytoplasmic transcripts. At the ER membrane, CAP-seq enriched mRNAs encoding for secretory pathway proteins, which is consistent with the model of co-translational protein targeting. At the outer mitochondrial membrane (OMM), we discovered 30 mRNAs encoding for protein subunits in the oxidative phosphorylation pathway (OXPHOS), supporting a model of local protein synthesis coupled with mitochondrial import. Our dataset also enriched 55 mRNAs encoding for cytoplasmic ribosomal subunits, which raises a question regarding the functional implications of such a spatial arrangement.

Results

Development of the CAP-seq technique. Nucleobases are known to be oxidized by a variety of reactive oxygen species (ROS), including hydroxyl radical and singlet oxygen[22]. KillerRed[23] and miniSOG[24] are genetically encoded photosensitizers that generate ROS (such as singlet oxygen[24] and superoxide[25,26]) on visible light illumination to achieve photo-ablation of neighboring protein

[1]School of Life Sciences, Tsinghua University, Beijing, China. [2]College of Chemistry and Molecular Engineering, Peking University, Beijing, China. [3]Synthetic and Functional Biomolecules Center, Peking University, Beijing, China. [4]Beijing National Laboratory for Molecular Sciences, Peking University, Beijing, China. [5]Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, Peking University, Beijing, China. [6]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [7]Peking-Tsinghua Center for Life Sciences, Beijing, China. [8]Peking-Tsinghua-NIBS Joint Graduate Program, Tsinghua University, Beijing, China. [9]Department of Chemical Biology, College of Chemistry, Nankai University, Tianjin, China. [10]Chinese Institute for Brain Research, Beijing, China. [11]PKU-IDG/McGovern Institute for Brain Research, Beijing, China. [12]These authors contributed equally: Pengchong Wang, Wei Tang, Zeyao Li. *e-mail: jianbinwang@mail.tsinghua.edu.cn; zoupeng@pku.edu.cn

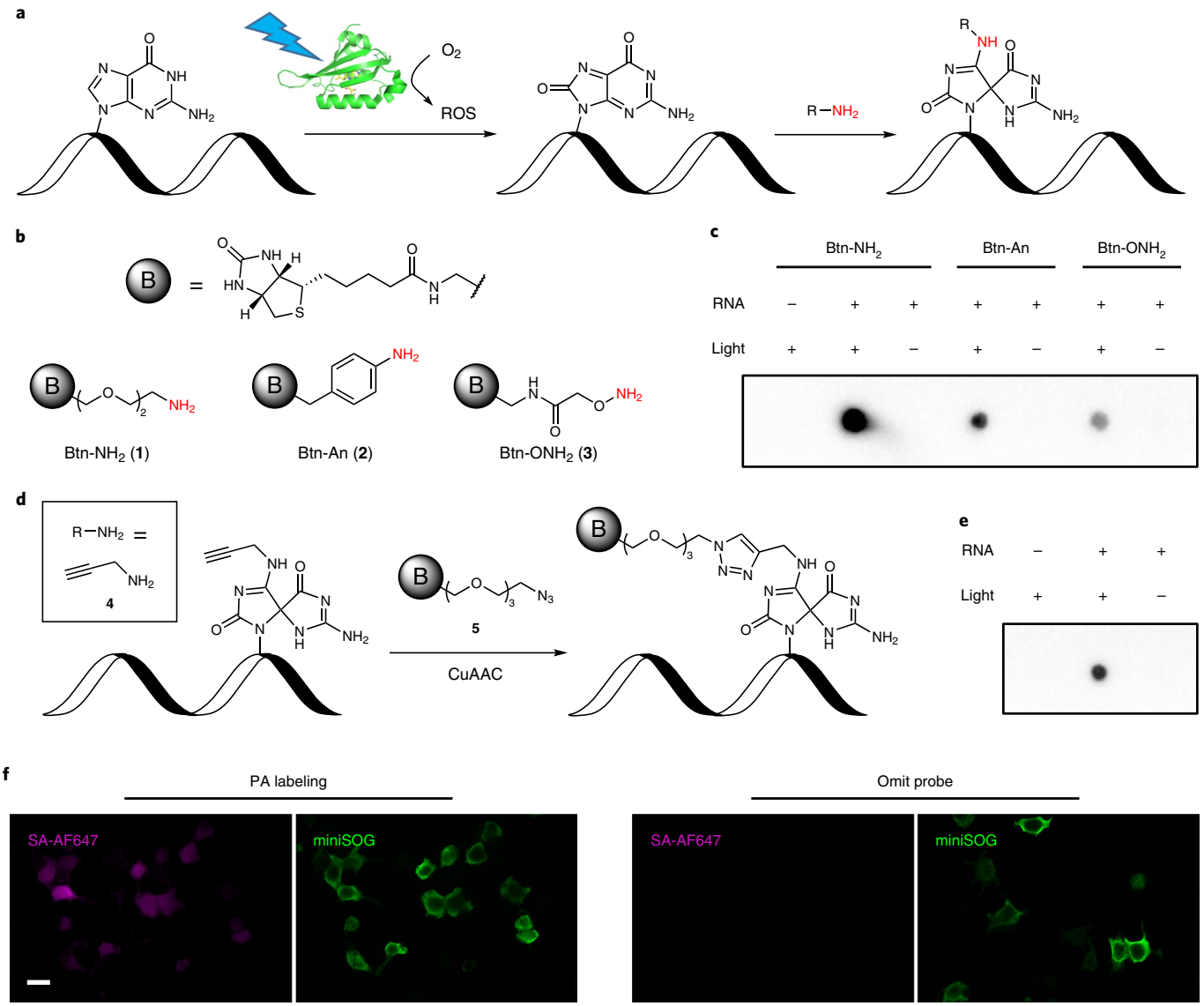

Fig. 1 | Development and characterization of CAP-seq method. a, Schematic of miniSOG-mediated RNA labeling. On visible light illumination, miniSOG generates ROS in the solution, which reacts with guanine nucleobase in RNA. The photo-oxidation intermediate could be intercepted by amine probes (R-NH_2) to form a covalent adduct. **b**, The chemical structure of biotin-conjugated amine probes. **c**, Representative streptavidin dot blot analysis of miniSOG-mediated RNA biotinylation with probes in **b**. **d**, Copper-assisted click reaction to introduce biotin to alkyne-conjugated RNA. **e**, Representative streptavidin dot blot analysis of biotinylated RNA from **d**. **f**, Representative immunofluorescence images of HEK293T cells expressing miniSOG in the cytoplasm. Left, cells are labeled with PA (**4**), derivatized with biotin-azide and stained with streptavidin-conjugated AlexaFluor647 (SA-AF647). Right, negative control experiment omitting the PA probe. Scale bar is 20 μm. Dot blot and immunofluorescence imaging experiments were independently repeated twice with similar results.

targets[23,27]. In this study, we sought to repurpose these photosensitizers to achieve proximity labeling of RNA in live cells. We reason that, due to the short lifetime and limited diffusion radius of ROS (<0.6 μs and 70 nm in cells for singlet oxygen[28]), such labeling reaction occurs only proximal to miniSOG and the tagged RNA could be subsequently enriched and identified through high-throughput sequencing.

We started by testing variants of miniSOG and KillerRed for their capability to oxidize nucleobases in vitro. High-performance liquid chromatography and tandem mass spectrometry analysis revealed that guanosine was converted into photo-oxidation products, imidazolone and spiroiminodihydantoin, in the presence of miniSOG and blue light illumination (Supplementary Fig. 1). Under these conditions, approximately half of the guanosine starting material was consumed over 20-min illumination, while no photo-conversion was observed for other nucleobases (Supplementary Fig. 2) or with KillerRed variants as the photosensitizer (Supplementary Fig. 3), presumably due to differences in ROS quantum yields[26] and nucleobase redox potentials[29].

Our next aim was to enrich photo-oxidized RNA. Guanosine is known to crosslink with amines during oxidative DNA damage[30,31]. We therefore sought to employ photosensitized guanosine oxidation to install biotin onto RNA (Fig. 1a). We set out to test a panel of biotin-conjugated amine probes, including alkylamine (Btn-NH_2, **1**), aniline (Btn-An, **2**) and alkoxyamine (Btn-ONH_2, **3**) (Fig. 1b). Using a streptavidin dot blot assay as the readout for biotinylation, we optimized the photosensitization (Supplementary Fig. 4), click reaction (Supplementary Fig. 5) and identified alkylamine

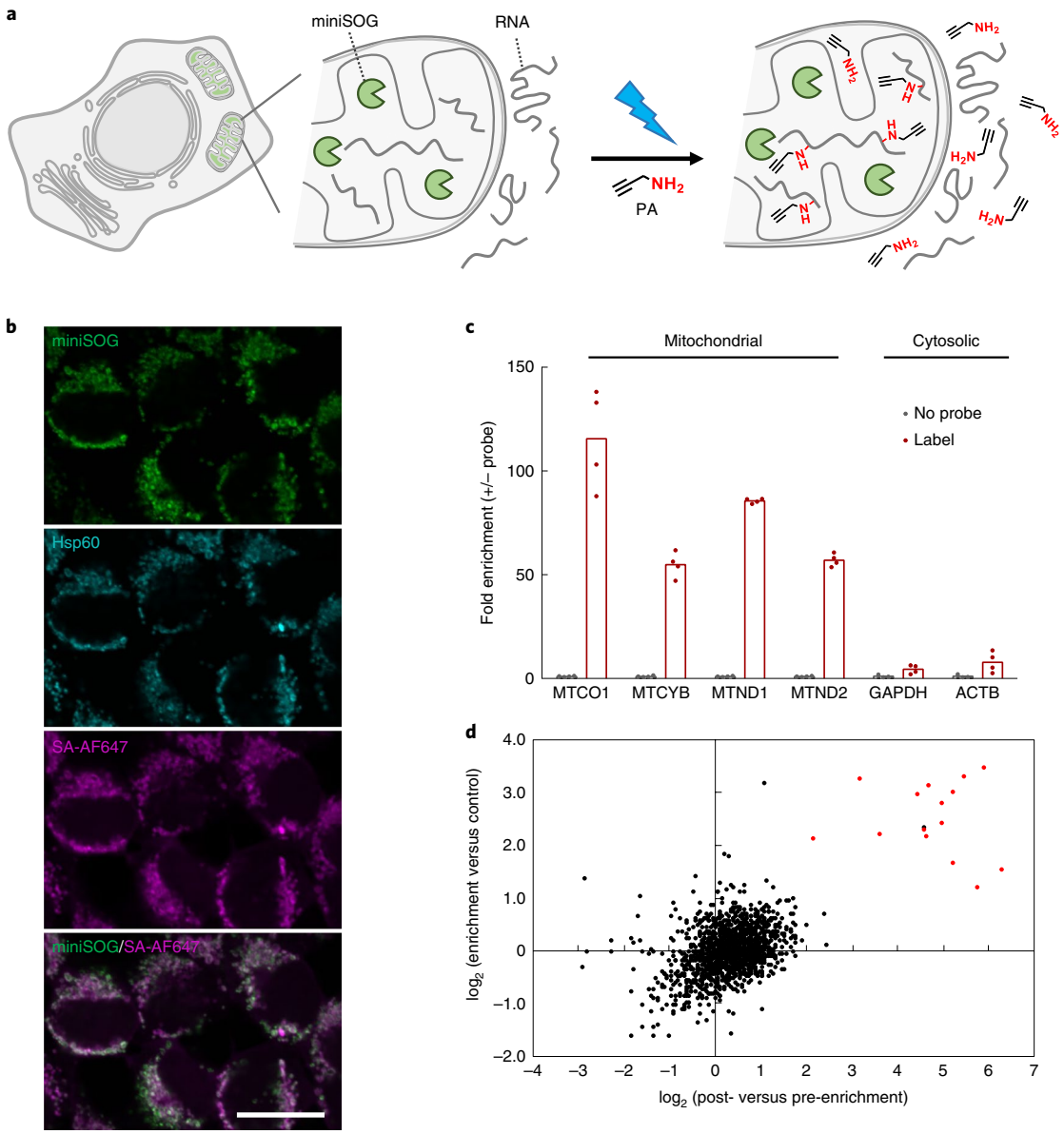

Fig. 2 | CAP-seq reveals subcellular transcriptome in the mitochondrial matrix. a, Scheme for miniSOG-mediated RNA labeling. miniSOG is genetically targeted to the mitochondrial matrix via fusion to the N-terminal 23 amino acid residues of human COX4 (mito-miniSOG). Cells are incubated with a PA probe and exposed to blue LED illumination. **b**, Representative images of HEK293T cells stably expressing mito-miniSOG (green). Hsp60 is a mitochondrial marker (cyan). Streptavidin-conjugated AlexaFluor647 (SA-AF647) staining shows biotinylation (magenta). Scale bar is 20 µm. This experiment was independently repeated five times with similar results. **c**, RT–qPCR analysis of enriched RNA. *MTCO1*, *MTCYB*, *MTND1* and *MTND2* are MT-mRNAs; *GAPDH* and *ACTB* are cytosolic mRNA markers. Relative abundance is calculated as the ratio of enrichment relative to the negative control omitting the PA probe. Data are the mean of four technical replicates ± 1 s.d. **d**, Scatter plot showing the enrichment ratios for each gene. Horizontal axis shows the FPKM ratio of post- versus pre-enrichment. Vertical axis shows the FPKM ratio of enrichment versus control (omitting probe). Both axes are plotted on logarithm units. The 15 red dots are MT-mRNAs and MT-rRNAs. The black dot in the enriched population is a mitochondrial pseudogene.

as the most reactive probe (Fig. 1c). Liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) analysis confirmed guanosine adduct formation (Supplementary Fig. 6). Dot blot analysis revealed that biotinylation occurred less than once per ~640 nucleobases on average (Supplementary Fig. 7) and preferentially targeted single-stranded nucleotide over double-stranded structure (Supplementary Fig. 8). A quantitative PCR with reverse transcription (RT–qPCR) assay verified that these biotinylated RNA molecules could be reverse transcribed and enriched with streptavidin-coated beads (Supplementary Fig. 9). Finally, to improve membrane permeability of the probe, we replaced the biotin moiety with terminal alkyne as a functional handle. Following photo-oxidative conjugation and RNA extraction, biotin was introduced via click chemistry (Fig. 1d,e and Supplementary Fig. 10). As shown in Fig. 1f, miniSOG successfully labeled cellular targets with propargyl amine (PA, **4**) (Fig. 1d) probe. Taken together, we concluded from these data that miniSOG and PA were the optimal combination for CAP-seq, which we used throughout the following experiments.

To determine the type of ROS responsible for the observed RNA labeling, we added various quenchers to the reaction mixture. Whereas miniSOG-mediated RNA labeling was greatly reduced in the presence of NaN_3 and trolox, which are known to quench singlet

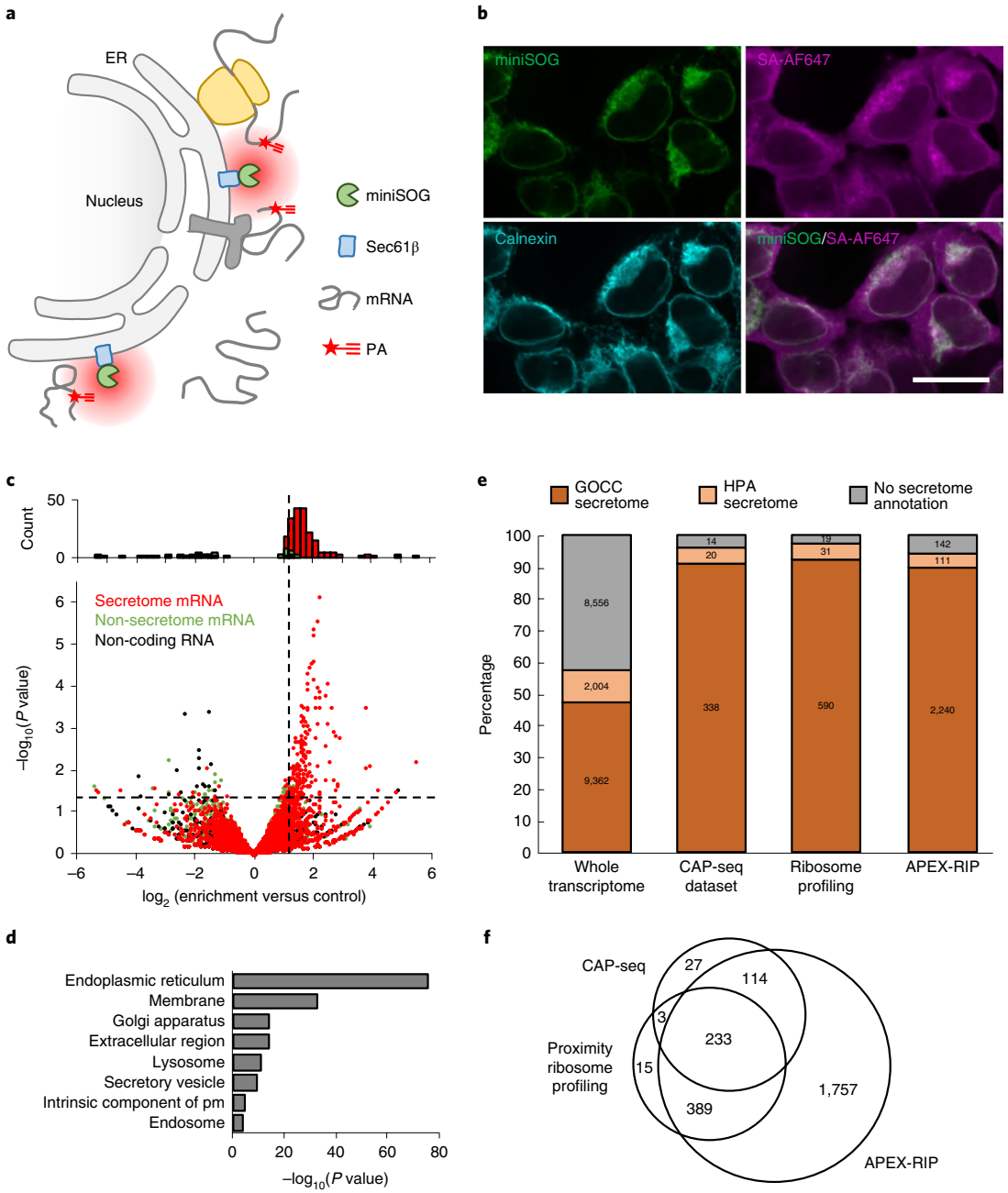

Fig. 3 | Mapping of the ER-proximal transcriptome. a, Schematic of CAP-seq at the ER membrane. miniSOG is targeted to the cytoplasmic side of the ER membrane via fusion with Sec61β. **b**, Representative immunofluorescence images showing miniSOG expression (green), the biotinylation signal (SA-AF647, magenta), ER marker calnexin (cyan) and the overlay. Scale bar is 20 μm. This experiment was independently repeated four times with similar results. **c**, Volcano plot of the ERM CAP-seq dataset. Red, green and black dots represent secretome mRNAs, nonsecretome mRNAs and noncoding RNAs, respectively. Horizontal dashed line indicates $P = 0.05$. Vertical dashed line indicates the cutoff of $\log_2$ (enrichment versus control) = 1.12, as determined from ROC analysis. Upper panel, histograms showing the distribution of RNAs with $P < 0.05$, whose color code is the same as dots in the volcano plot. *P* values were calculated using the Wald test in DESeq2 software. **d**, GOCC analysis of 372 enriched mRNA in the ERM CAP-seq dataset. **e**, Comparison of the secretome specificity of CAP-seq and other transcriptome profiling techniques. Only protein-coding mRNAs are included in the analysis. The number of transcripts in each condition is indicated in the columns. Secretome annotation is generated from GOCC, the HPA and MitoCarta databases (see Methods). **f**, Venn diagram comparing the datasets of ERM CAP-seq, proximity ribosome profiling and APEX-RIP.

oxygen[27,32], it was relatively insensitive toward the hydroxyl radical scavenger, mannitol[27] (Supplementary Fig. 11). The presence of 78% (v/v) D_2O, known to stabilize singlet oxygen, enhanced the labeling signal. These data strongly indicate singlet oxygen as the major ROS that gave rise to RNA biotinylation. The pH-activity profile of miniSOG-mediated RNA labeling exhibited a sharp decrease at pH 4.5, suggesting that its application should avoid acidic cellular environment such as the lysosomal lumen. Dot blot analysis also revealed that RNA labeling was insensitive to Ca^{2+} but was reduced in the presence of glutathione and amino acids at physiologically relevant concentrations, or proteins at 30 mg ml^{-1} (Supplementary Fig. 12). Consistent with this observation, we found that the yield

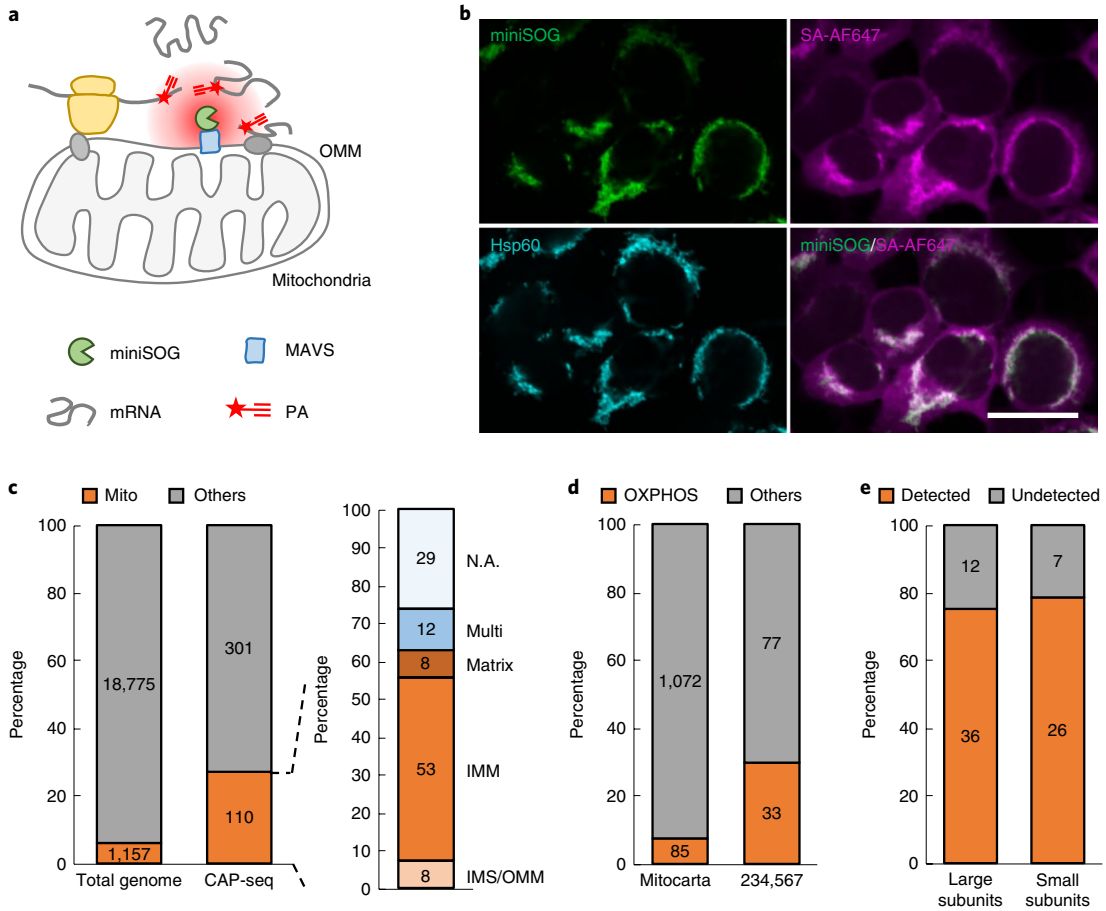

Fig. 4 | Mapping local transcriptome at the OMM. a, Schematic of CAP-seq at the OMM. miniSOG is targeted to the OMM via fusion with the C terminal 31 amino acid residues of MAVS. **b**, Representative immunofluorescence images of miniSOG expression (green), biotinylation signal (SA-AF647, magenta), mitochondrial marker Hsp60 (cyan) and the overlay. Scale bar is 20 μm. This experiment was independently repeated four times with similar results. **c**, Enrichment of mitochondrial genes in the OMM CAP-seq dataset (left). The sub-mitochondrial distribution of enriched mRNAs is shown on the right. Matrix, mitochondrial matrix; Multi, multiple sub-mitochondrial locations as annotated in GOCC. N.A., no submitochondrial annotations available. **d**, Enrichment of OXPHOS genes in the OMM CAP-seq dataset. **e**, Coverage of cytosolic ribosomal genes in the OMM CAP-seq dataset, including both large and small subunits.

of miniSOG-mediated RNA labeling in vivo was approximately 1,500-times lower than in vitro, occurring approximately once per 9.3×10^5 nucleobases on average (Supplementary Fig. 13). Decreased yield was likely due to quenching of singlet oxygen in the cellular environment and competition of PA with amine-containing biomolecules. In addition to labeling RNA, we also noted that miniSOG could mediate the photo-oxidation of DNA and proteins (Supplementary Fig. 14). Therefore, care must be taken to remove contaminating DNA during sample preparation.

We next turned to test the spatial resolution and sensitivity of miniSOG-mediated RNA labeling in live cells. We chose the mitochondrial matrix as a model (Fig. 2a), which has a well-established transcriptome, containing 13 mRNAs (MT-mRNAs), two ribosomal RNAs (MT-rRNAs) and 22 transfer RNAs[11]. We targeted miniSOG to the mitochondrial matrix of human embryonic kidney 293T (HEK293T) cells (Supplementary Fig. 15), exogenously supplied cells with PA in the culturing medium and initiated the reaction with blue light illumination at 24 mW cm^{-2}. It is not necessary to strictly avoid regular laboratory light before the labeling step. The reaction was terminated after 20 min and the cells were either fixed for imaging analysis or lysed to extract RNA content. Immunostaining data showed that biotinylation signal highly colocalized with miniSOG (Fig. 2b). Following cell lysis and RNA extraction, the sample was treated with DNase to remove residual DNA. Purified RNA was reacted with biotin-azide in the presence of Cu(I) catalyst, enriched with streptavidin-coated beads and analyzed by RT–qPCR or RNA-seq (Supplementary Fig. 16). Labeling signal was completely removed when purified sample was treated with RNase A/T1 (Supplementary Fig. 17). We optimized the illumination time (20 min), probe concentration (10 mM) and reverse transcriptase (ProtoScript II) (Supplementary Figs. 18–20). Under this condition, mitochondrial RNAs were highly enriched (Fig. 2c) compared to the negative control omitting the PA probe, whereas cytosolic RNA markers such as *ACTB* and *GAPDH* were not enriched. Successful labeling of mitochondrial RNAs indicated good membrane permeability of the PA probe, as it needed to pass through both the plasma membrane and the mitochondrial inner membrane to reach its target.

To analyze RNA-seq data, we started by verifying that miniSOG expression, probe incubation and photo-oxidation did not alter the landscape of whole cell transcriptome (Supplementary Figs. 21 and 22). We then compared the FPKM (fragments per kilobase per million mapped reads) values for each gene before versus after affinity purification, and in the presence versus absence of PA probe. A plot of these ratios revealed enrichment of 16 genes across duplicate experiments (Fig. 2d), including all 13 MT-mRNAs, both MT-rRNAs and one mitochondrial pseudogene MTATP6P1 (Source Data Fig. 2). Consistent with RT–qPCR results, none of the

cytosolic RNAs were substantially enriched in our dataset. Together, the above data demonstrated that CAP-seq method is capable of labeling cellular transcriptome in a proximity-dependent manner.

Enrichment of secretome mRNA at the endoplasmic reticulum membrane (ERM). We next turned to another important subcellular compartment, the ERM (Fig. 3a). Unlike the mitochondrial matrix, which is enclosed by two layers of lipid membranes that help prevent the leakage of ROS, the ERM represents more 'open' space and is thus considered a more challenging and rigorous test for proximity-labeling techniques[20]. ERM is also an interesting target because it exemplifies local protein translation—secretory pathway proteins are initially synthesized at the ERM. We therefore expected to enrich mRNAs encoding for the secretome. Indeed, Jan et al. applied proximity ribosome profiling technique to investigate the local translation at the ERM in mammalian cells[33]. More recently, Kaewsapsak et al. combined APEX-mediated protein labeling and protein–RNA crosslinking (APEX-RIP)[20] to obtain a list of ERM proximal RNAs. These provided a good reference for comparison with our dataset.

We created a HEK293T cell line stably expressing miniSOG targeted to the ERM via fusion to the N terminus of ER translocon Sec61β subunit (Supplementary Fig. 23). Immunostaining data showed that biotinylation signal had a high degree of overlap with miniSOG targeting, although it appeared to be more diffusive (Fig. 3b), presumably due to RNA translocation before cellular fixation. We performed two independent replicate experiments and selected polyadenylated RNA for sequencing analysis (Fig. 3c). Using receiver operating characteristic (ROC) analysis[34], we determined the threshold of the $\log_2$ (enrichment versus control) to be 1.12 (Supplementary Fig. 24). This yielded a list of 377 RNAs, including 372 mRNAs (99%), three pseudogenes, one antisense RNA and one long noncoding RNA (Source Data Fig. 3). Gene Ontology analysis revealed that these mRNAs primarily encoded secretory pathway proteins (Fig. 3d), consistent with the model of local translation at the ERM. As shown in the case of *TIMP1*, sequencing reads mapped almost uniformly to the transcript (Supplementary Fig. 25).

To quantitatively assess the specificity of ERM CAP-seq dataset, we defined a list of secretome genes based on annotations in the Gene Ontology Cellular Component (GOCC) and Human Protein Atlas (HPA) databases, and 358 out of 372 enriched mRNAs (96.2%) belong to the secretome. This level of specificity is comparable to ribosome profiling (97%) and APEX-RIP (94%) (Fig. 3e). Also, 63 and 92% of the transcripts identified by ERM CAP-seq (236 and 348 out of 377) overlapped with the proximity ribosome profiling dataset[33] and the APEX-RIP dataset[20], respectively (Fig. 3f). A total of 26 RNAs were newly discovered in the current study (Source Data Fig. 3). The above analysis showed that CAP-seq approach offered exceptional spatial specificity in 'open' subcellular space, and could complement existing methods.

Identification of RNAs proximal to the OMM. To further demonstrate the power of CAP-seq technique, we investigated the local transcriptome near the OMM (Fig. 4a). Except for 13 MT-mRNA

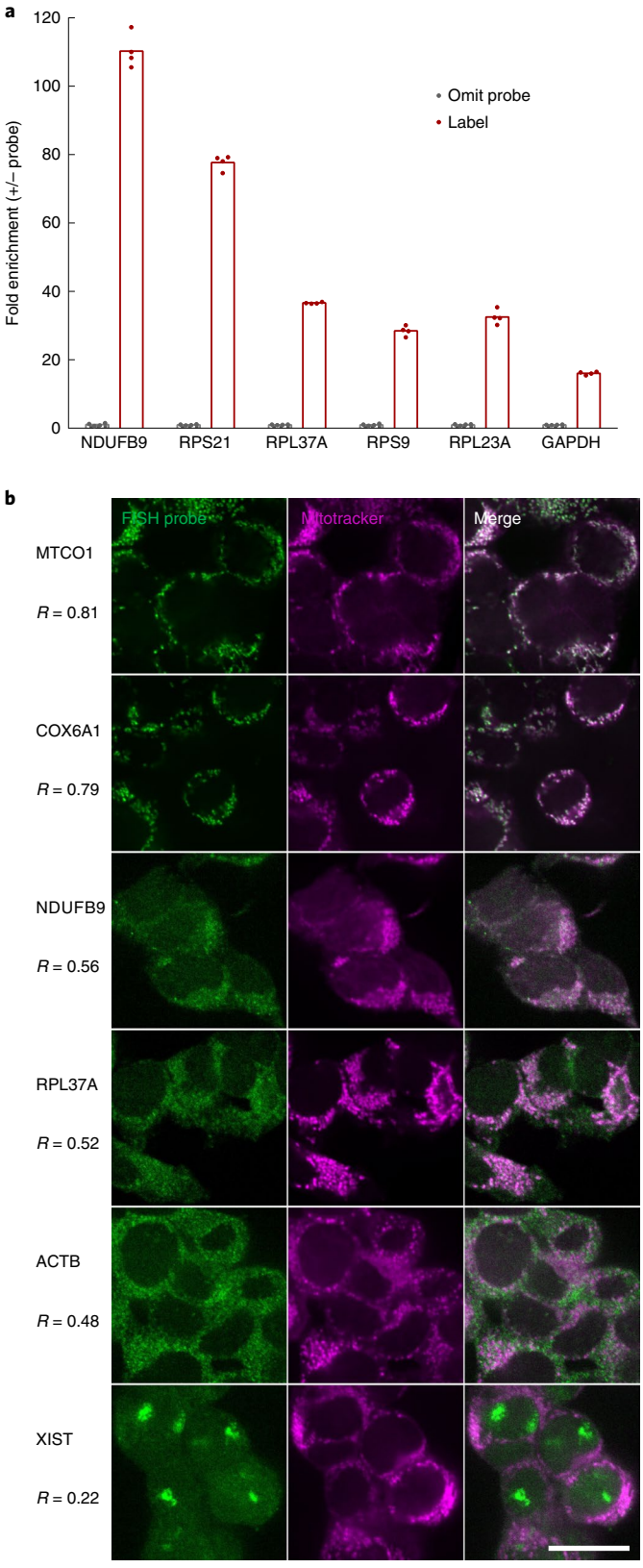

Fig. 5 | Validation of RNAs identified in OMM CAP-seq. a, RT–qPCR analysis of RNAs enriched in OMM CAP-seq dataset. *NDUFB9* is an OXPHOS mRNA. *RPS21*, *RPL37A*, *RPS9* and *RPL23A* are ribosomal protein-coding mRNA. *GAPDH* is a cytosolic mRNA marker. Fold enrichment is the calculated as the ratio of enrichment in the labeled sample versus the negative control omitting the PA probe. Data are the mean of four technical replicates ± 1 s.d. **b**, Representative RNA FISH images of transcripts identified by OMM CAP-seq. Left, fluorescence images of FISH probes targeted to RNAs (green). Middle, mitotracker staining of mitochondria (magenta). Right, merged images of FISH and mitotracker. *COX6A1*, *NDUFB9* and *RPL37A* are mRNAs enriched in OMM CAP-seq. *MTCO1*, a mitochondrial genome-encoded mRNA, is the positive marker. *ACTB* is a cytosolic mRNA marker. *XIST* is a nuclear-localized lncRNA marker. *R* is the Pearson's correlation coefficient of FISH image and mitotracker staining (see Methods). Scale bar is 20 μm. This experiment was independently repeated four times with similar results.

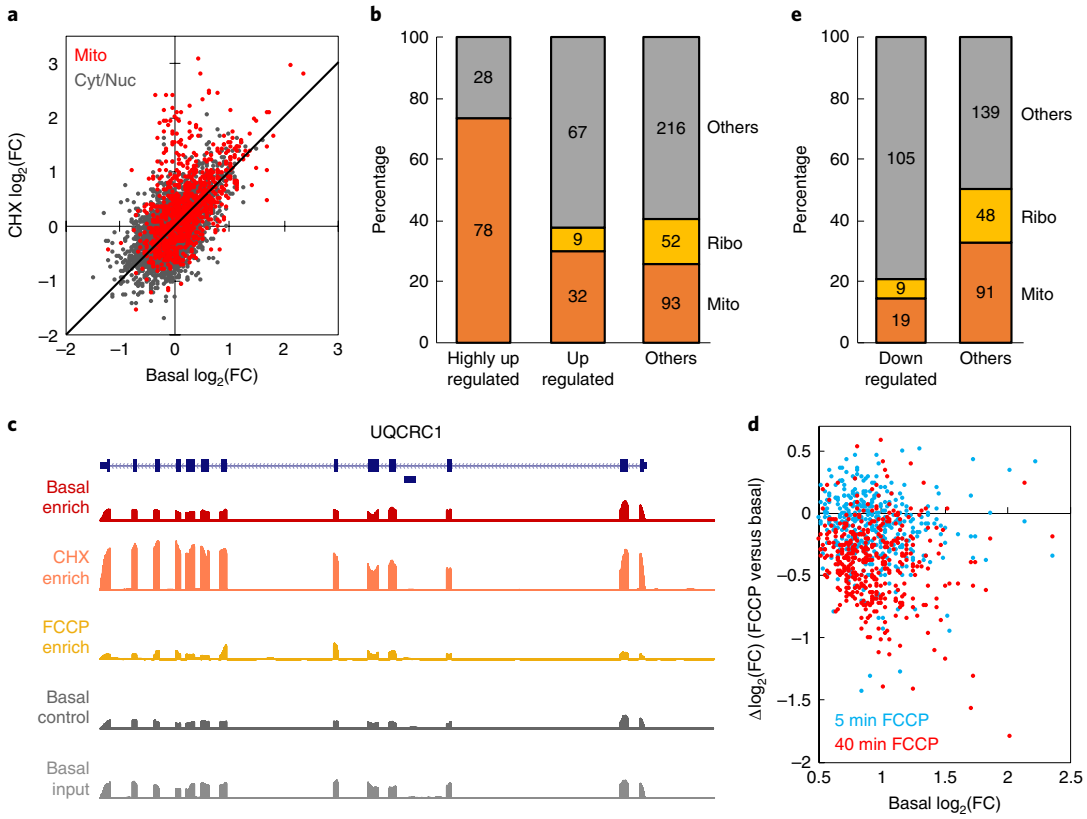

Fig. 6 | OMM CAP-seq with drug perturbation. a, Scatter plot of OMM CAP-seq $\log_2$(FC) comparing the CHX-treated sample (*y* axis) versus the basal level (*x* axis). Red dots represent mRNA encoding for mitochondrial proteins annotated in MitoCarta. Gray dots represent mRNAs encoding for cytosolic and nuclear proteins, excluding mitochondrial or secretory proteins. **b**, Breakdown of mRNAs encoding for mitochondrial proteins and ribosomal subunits in three enriched fractions as defined by $\Delta\log_2$(FC) (CHX versus basal). Highly upregulated: $\Delta\log_2(FC) > 0.8$; upregulated: $\Delta\log_2$(FC) between 0.5 and 0.8; others: $\Delta\log_2(FC) < 0.5$. **c**, Genome tracks of *UQCRC1*, an OXPHOS gene encoding for a protein subunit in mitochondrial complex III, enriched at the basal level, on treatment with CHX and FCCP, respectively. **d**, Scatter plot showing $\Delta\log_2$(FC) (FCCP versus basal) after 5-min (blue) or 40-min (red) FCCP treatment (*y* axis), against OMM CAP-seq $\log_2$(FC) at the basal level (*x* axis). **e**, Breakdown of mRNAs encoding for mitochondrial proteins and ribosomal subunits in FCCP-treated OMM CAP-seq dataset. Downregulated, $\Delta\log_2$(FC) (FCCP versus basal) < -0.5; others, $\Delta\log_2(FC) > -0.5$.

genes, the rest of the mitochondrial proteome is encoded by the nuclear genome. Some of these mRNAs are known to be localized to the OMM such that protein import occurs during or immediately after translation[35,36]. We stably expressed OMM-targeted miniSOG (via N-terminal fusion with MAVS) in HEK293T cells (Supplementary Fig. 26), and verified the spatial specificity of RNA labeling with immunostaining (Fig. 4b). For RNA-seq, we performed five independent replicate experiments and obtained a list of 472 transcripts ($P < 0.01$, Source Data Fig. 4), including 411 mRNAs (87%). By comparing our dataset to the established inventory of human mitochondrial genes, MitoCarta v.2.0 (ref. [37]), we found that 110 mRNAs (27%) encoded mitochondrial proteins (Fig. 4c). As a reference, MitoCarta genes only accounted for 5.8% of the human genome. Nearly half of these mRNAs (53 out of 110) encoded for inner mitochondrial membrane (IMM) proteins (Fig. 4c), including 33 out of 85 protein subunits in the OXPHOS pathway (Fig. 4d). Supplementary Fig. 25 showed the genome track of *NDUFB9*, a highly enriched OXPHOS gene. Our dataset strongly indicated local translation of these membrane proteins at the OMM. This may be advantageous because protein synthesis near the mitochondrial membrane might avoid exposing hydrophobic transmembrane domains to the cytoplasm, thus reducing the risk of protein aggregation or aberrant sorting to other membranes. Consistent with our observation in human cells, previous ribosome profiling data in yeast also discovered IMM components in the enriched gene list[35].

For the remaining 301 enriched mRNAs, GOCC analysis surprisingly revealed cytosolic ribosomal components (Supplementary Fig. 27). The human cytosolic ribosome contains 48 and 33 proteins in the large and small subunits, respectively[38,39]. Among these, OMM CAP-seq identified 75% of mRNAs (36 out of 48) encoding for large subunit proteins and 79% (26 out of 33) for small subunit proteins (Fig. 4e). Genome tracks of two representative cytosolic ribosomal genes, *RPS21* and *RPL23A*, are shown in Supplementary Fig. 25. Notably, none of these was enriched in the ERM dataset. In addition to mRNAs, OMM CAP-seq also captured other class of RNA species, including 55 pseudogenes (Supplementary Fig. 28). Of these pseudogenes, 40% (22 out of 55) were related to the cytosolic ribosome (Source Data Fig. 4).

We validated highly enriched mRNAs in the OMM CAP-seq dataset with RT–qPCR and RNA FISH imaging. RT–qPCR analysis showed that the nuclear-encoded mitochondrial mRNA *NDUFB9*, which was enriched 5.1-fold in OMM CAP-seq, was enriched 7.9-fold relative to the cytosolic marker *GAPDH* (Fig. 5a). Similarly, mRNAs encoding for proteins in both large (*RPL23A* and *RPL37A*) and small (*RPS9* and *RPS21*) ribosomal subunits were enriched 1.8- to 4.8-fold relative to *GAPDH* (Fig. 5a). For RNA FISH imaging, we stained mitochondria with Mitotracker and chose the mitochondrial-encoded mRNA *MTCO1* as a positive marker, whose FISH signal colocalized with the Mitotracker staining pattern in HEK293T cells ($r_{Pearson} = 0.81$, Fig. 5b). For the negative marker, we chose the nuclear-localized lncRNA XIST, whose FISH signal exhibited poor

overlap with Mitotracker ($r_{Pearson}$ = 0.22, Fig. 5b). We went on to analyze nuclear-encoded OXPHOS mRNAs (*COX6A1* and *NDUFB9*) and ribosomal protein-coding mRNA *RPL37A*. Image correlation coefficients of these mRNAs with respect to the Mitotracker fell between those of the positive and negative markers ($r_{Pearson}$ from 0.52 to 0.79, Fig. 5b). For comparison, the cytosolic marker mRNA *ACTB* had a $r_{Pearson}$ of 0.48 (Fig. 5b).

We further compared our ERM and OMM CAP-seq datasets to check for RNAs with potential dual localizations. The plot in Supplementary Fig. 29 shows two distinct RNA populations, confirming that each of these experiments enriched a unique set of transcripts. The overlap between ERM and OMM datasets yielded 95 RNAs, nearly all of which (94/95) were previously annotated as secretome-encoding (Source Data Fig. 4). These mRNAs may be localized at the mitochondria-ER contact sites.

Two possible mechanisms may exist to explain the mitochondrial targeting of RNA: RNA could either actively and directly associate with RNA-binding proteins on the OMM, or be passively targeted in the form of ribosome/mRNA/nascent peptide ternary complex during protein translation (Supplementary Fig. 30). To distinguish between these mechanisms, we employed cycloheximide (CHX), a protein synthesis inhibitor, to stabilize the ternary complex[33,35]. We treated HEK293T cells with CHX before OMM CAP-seq, and noticed a dramatic increase in RNA enrichment levels across duplicate experiments (Fig. 6a), particularly for nuclear-encoded mitochondrial mRNAs (Fig. 6b). We reasoned that CHX may cause the accumulation of ribosome-mediated RNA targeting to the OMM by slowing down ribosome recycling. Figure 6c shows the genome track of *UQCRC1*, whose enrichment was enhanced by 6.2-fold in the sequencing data. In contrast, OMM enrichment of ribosomal protein-coding RNAs only slightly increased. To quantify the effect of CHX on OMM enrichment, we calculated the differences of mRNA enrichment levels measured in CHX-treated samples versus untreated control (defined as $\Delta\log_2$(fold change (FC))), and applied 0.5 as the cutoff. This yielded a list of 214 CHX-sensitive mRNA, with 110 mitochondrial RNA (51%) and nine ribosomal protein-coding RNA (4%) (Source Data Fig. 6a). Notably, the percentage of mitochondrial RNA was 1.9-fold higher than the basal level (51 versus 27%), whereas that of ribosomal protein-coding RNA was 3.6-fold lower (4 versus 15%).

How does mitochondrial activity affect RNA targeting to the OMM? Carbonyl cyanide *p*-trifluoromethoxyphenylhydrazone (FCCP) is an ionophore that depolarizes the mitochondrial membrane potential within minutes. We applied 10 µM FCCP to HEK293T cells for 5–40 min and measured RNA enrichments levels with duplicate CAP-seq experiments. Whereas treatment for 5 min did not appreciably change enrichment relative to the basal level, longer incubation (40 min) caused a dramatic decrease in enrichment levels (Fig. 6d). We chose $\Delta\log_2(FC) < -0.5$ as the filter to define FCCP-sensitive transcripts, and obtained a list of 133 RNA, with 19 mitochondrial RNA (14%), nine ribosomal protein-coding RNA (7%) and 95 secretome mRNA (71%) (Fig. 6e and Source Data Fig. 6d). Notably, the lists of CHX-sensitive and FCCP-sensitive RNAs were almost mutually exclusive to each, with only four overlapping RNAs. (Supplementary Fig. 31). Consistent with this observation, the heat map in Supplementary Fig. 32 shows that CHX-sensitive RNAs were not enriched in the ERM dataset and vice versa. Taken together, we concluded from these data that mitochondrial membrane potential could promote the association of secretome mRNA with OMM, but was dispensable for the ribosome-mediated targeting pathway.

Recently, eosin-mediated photo-oxidation of RNA and proteins were combined with Halotag targeting to label subcellular transcriptome in the nucleus and the cytoplasm[40,41]. The application of small molecule dyes may cause high background due to nonspecific binding, which could reduce the spatial specificity. In comparison, CAP-seq is fully genetically encoded. We compared eosin- and miniSOG-mediated RNA labeling at the OMM with RT–qPCR, which showed that CAP-seq approach yielded higher levels of RNA enrichment (Supplementary Fig. 33).

Discussion

In this study, we report a proximity-dependent RNA labeling method, CAP-seq. This simple approach allows for unbiased, transcriptome-wide profiling of RNA spatial distribution in cells, without requirements of cellular fixation, sophisticated instruments or previous knowledge of RNA sequences. Using the mitochondrial matrix as a model system, we demonstrated the high spatial specificity and good depth of coverage of CAP-seq in live cells. We applied CAP-seq technique to investigate the local transcriptome near the surface of the ERM and OMM.

APEX has been used for proximity-dependent proteomic labeling[19]. By combining APEX-mediated protein biotinylation with formaldehyde-mediated or photo-induced protein–RNA crosslinking, two new technologies, called APEX-RIP and Proximity-CLIP, were recently developed to profile local transcriptome[20,21]. In addition to capturing RNA, Proximity-CLIP has the advantage of revealing RNA-interacting proteins[21], which is not possible with CAP-seq. For membrane-free subcellular regions, such as the ERM, APEX-RIP has exhibited poorer spatial resolution when APEX2 was targeted to the cytoplasmic face of ERM (for example, positive markers encoding for ER membrane proteins *EMC10* and *SSR2* were not enriched), presumably due to perturbations from formaldehyde treatment and protein diffusion during APEX labeling[20]. In comparison, the CAP-seq technique avoids problems associated with formaldehyde crosslinking and enables the investigation of local transcriptome in nonmembrane-bound regions such as the OMM. Our ERM CAP-seq dataset enriched both *EMC10* (3.2-fold) and *SSR2* (2.6-fold).

Notably, the recent development of APEX-seq, which is done in parallel to this study, has also reported promising spatial specificity by avoiding the crosslinking step[42]. Because CAP-seq and APEX-seq employ distinct mechanisms of RNA labeling, they may differ in their spatial resolution and target specificity and could thus complement each other in profiling the spatial transcriptome. APEX-seq has exhibited superior temporal resolution over CAP-seq (1 versus 20 min), which is beneficial when studying fast biological processes such as signaling cascades[42]. The fast labeling kinetics might also facilitate the detection of short-lived RNA transcripts, whose half-lives might be as few as 10 min (ref. [43]). For RNA molecules with turn-over rates faster than the temporal resolution, a substantial portion would be degraded during labeling, which may reduce the efficiency of downstream affinity capture and complicate the analysis of sequencing results.

At the ERM, our method enriched 372 mRNA with 96% specificity toward secretome-encoding genes. Because miniSOG was targeted to the ERM facing cytosol, mRNAs encoding for nonsecretory proteins should also be accessible to miniSOG, although at a longer distance. The absence of abundant cytosolic RNAs indicated that the 'action radius' of this approach was limited to only a few nanometers. Compared to existing datasets obtained by ribosome profiling[33] and APEX-RIP[20], our study discovered 26 new RNAs that are proximal to the ERM, including 22 mRNAs and four noncoding RNAs. Then, 11 of the 22 mRNAs were annotated to be secretory RNAs, such as *IFNGR2* and *ENHO*. *IFNGR2* encodes interferon gamma receptor 2, which associates with IFNGR1 to form the receptor for the cytokine interferon gamma. IFNGR2 is an Federal Drug Administration approved drug target and its mutation is associated with Mendelian susceptibility to mycobacterial disease[44]. *ENHO* encodes adropin, a secreted protein that regulates glucose and lipid homeostasis[45].

Our OMM CAP-seq dataset could potentially yield new knowledge of RNA targeting and function. Among 411 mRNAs enriched

in this study, 27% (110 transcripts) encode for mitochondrial proteins. This observation is consistent with previous report of local protein translation at the OMM. Results from our CHX-sensitivity assay further supported the model of ribosome-mediated OMM targeting. Our data also revealed 62 mRNAs encoding for the cytosolic ribosomal proteins, and we verified that their OMM targeting was insensitive to CHX or FCCP. Could these RNAs serve noncoding roles such as scaffolding? Or could they mediate the coordination of mitochondrial and cytosolic translation programs? Future efforts to analyze their local translation (for example, via subcellular ribosome profiling) and to identify their binding proteins (for example, via PAR-CLIP) might help answer these questions.

In the OMM CAP-seq dataset, we also noticed the enrichment of several redox enzyme genes: GPx1, GPx4 and SOD1. GPx1 and GPx4 are selenoperoxidases in the glutathione peroxidase family, which catalyze the clearance of cellular peroxide[46]. SOD1 is a free radical scavenger that catalyzes the dismutation of superoxide. Mutations in SOD1 has been linked to pathogenesis of amyotrophic lateral sclerosis[47]. Because mitochondria is a major source of peroxide and superoxide, the mitochondrial targeting of ROS scavengers could play important roles in protecting cells. Indeed, GPx4 inactivation was reported to induce notable morphological changes of mitochondria[48]. All three enzymes are found both in the cytoplasm and in the mitochondrial inter-membrane space (IMS), and previous studies have shown that their mitochondrial localization does not require mitochondrial targeting sequences[49,50]. Because enrichment of these three mRNAs were not sensitive toward CHX treatment (Source Data Fig. 6a), it is likely that their OMM targeting may not require intact ribosomes but rather depend on certain RNA-binding proteins. Other GPx family members, which resided outside the mitochondria[46], were not enriched by OMM CAP-seq.

We envision that CAP-seq method could be extended to investigate the RNA content in other genetically accessible subcellular regions, particularly those nonmembrane-bound organelles. The genetically encoded photosensitizer, miniSOG, is a small protein tag and has been demonstrated to function in many subcellular locations and organisms[24,27]. The high temporal resolution of CAP-seq method (20 min) would also allow for profiling changes in the local transcriptome as cells undergo physiological transitions (for example, during differentiation) or as cells respond to external stimuli (for example drug-induced stress).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41589-019-0368-5.

Received: 30 January 2019; Accepted: 19 August 2019;
Published online: 7 October 2019

References

1. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015).
2. Jung, H., Gkogkas, C. G., Sonenberg, N. & Holt, C. E. Remote control of gene function by local translation. *Cell* **157**, 26–40 (2014).
3. Sepulveda, G. et al. Co-translational protein targeting facilitates centrosomal recruitment of PCNT during centrosome maturation in vertebrates. *eLife* **7**, e34959 (2018).
4. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
5. Cioni, J. M., Koppers, M. & Holt, C. E. Molecular control of local translation in axon development and maintenance. *Curr. Opin. Neurobiol.* **51**, 86–94 (2018).
6. Glock, C., Heumuller, M. & Schuman, E. M. mRNA transport & local translation in neurons. *Curr. Opin. Neurobiol.* **45**, 169–177 (2017).
7. Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell. Biol.* **17**, 756–770 (2016).
8. Diehn, M., Bhattacharya, R., Botstein, D. & Brown, P. O. Genome-scale identification of membrane-associated human mRNAs. *PLoS Genet.* **2**, e11 (2006).
9. Chen, Q., Jagannathan, S., Reid, D. W., Zheng, T. & Nicchitta, C. V. Hierarchical regulation of mRNA partitioning between the cytoplasm and the endoplasmic reticulum of mammalian cells. *Mol. Biol. Cell* **22**, 2646–2658 (2011).
10. Marc, P. et al. Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep.* **3**, 159–164 (2002).
11. Mercer, T. R. et al. The human mitochondrial transcriptome. *Cell* **146**, 645–658 (2011).
12. Cajigas, I. J. et al. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* **74**, 453–466 (2012).
13. Minis, A. et al. Subcellular transcriptomics-dissection of the mRNA composition in the axonal compartment of sensory neurons. *Dev. Neurobiol.* **74**, 365–381 (2014).
14. Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat. Methods* **10**, 119–121 (2013).
15. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
16. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
17. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
18. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
19. Rhee, H. W. et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
20. Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L. & Ting, A. Y. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *eLife* **6**, e29224 (2017).
21. Benhalevy, D., Anastasakis, D. G. & Hafner, M. Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. *Nat. Methods* **15**, 1074–1082 (2018).
22. Fleming, A. M. & Burrows, C. J. Formation and processing of DNA damage substrates for the hNEIL enzymes. *Free Radic. Biol. Med.* **107**, 35–52 (2017).
23. Bulina, M. E. et al. A genetically encoded photosensitizer. *Nat. Biotechnol.* **24**, 95–99 (2006).
24. Shu, X. et al. A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* **9**, e1001041 (2011).
25. Westberg, M., Holmegaard, L., Pimenta, F. M., Etzerodt, M. & Ogilby, P. R. Rational design of an efficient, genetically encodable, protein-encased singlet oxygen photosensitizer. *J. Am. Chem. Soc.* **137**, 1632–1642 (2015).
26. Vegh, R. B. et al. Reactive oxygen species in photochemistry of the red fluorescent protein 'Killer Red'. *Chem. Commun.* **47**, 4887–4889 (2011).
27. Makhijani, K. et al. Precision optogenetic tool for selective single- and multiple-cell ablation in a live animal model system. *Cell. Chem. Biol.* **24**, 110–119 (2017).
28. Moan, J. On the diffusion length of singlet oxygen in cells and tissues. *J. Photoch. Photobio. B.* **6**, 343–347 (1990).
29. Morikawa, M. et al. Analysis of guanine oxidation products in double-stranded DNA and proposed guanine oxidation pathways in single-stranded, double-stranded or quadruplex DNA. *Biomolecules* **4**, 140–159 (2014).
30. Xu, X., Muller, J. G., Ye, Y. & Burrows, C. J. DNA-protein cross-links between guanine and lysine depend on the mechanism of oxidation for formation of C5 vs C8 guanosine adducts. *J. Am. Chem. Soc.* **130**, 703–709 (2008).
31. Ding, Y., Fleming, A. M. & Burrows, C. J. Sequencing the mouse genome for the oxidatively modified base 8-oxo-7,8-dihydroguanine by OG-Seq. *J. Am. Chem. Soc.* **139**, 2569–2572 (2017).
32. Bisby, R. H., Morgan, C. G., Hamblett, I. & Gorman, A. A. Quenching of singlet oxygen by Trolox C, ascorbate, and amino acids: effects of pH and temperature. *J. Phys. Chem. A.* **103**, 7454–7459 (1999).
33. Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521 (2014).
34. Linden, A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *J. Eval. Clin. Pract.* **12**, 132–139 (2006).
35. Williams, C. C., Jan, C. H. & Weissman, J. S. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science* **346**, 748–751 (2014).
36. Lesnik, C., Golani-Armon, A. & Arava, Y. Localized translation near the mitochondrial outer membrane: an update. *RNA Biol.* **12**, 801–809 (2015).

37. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).
38. Anger, A. M. et al. Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**, 80–85 (2013).
39. Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520**, 640–645 (2015).
40. Li, Y., Aggarwal, M. B., Nguyen, K., Ke, K. & Spitale, R. C. Assaying RNA localization in situ with spatially restricted nucleobase oxidation. *ACS Chem. Biol.* **12**, 2709–2714 (2017).
41. Li, Y., Aggarwal, M. B., Ke, K., Nguyen, K. & Spitale, R. C. Improved analysis of RNA localization by spatially restricted oxidation of RNA-protein complexes. *Biochemistry* **57**, 1577–1581 (2018).
42. Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell* **178**, 1–18 (2019).
43. Chen, C. Y., Ezzeddine, N., & Shyu, A. B. & Messenger, R. N. A. MehMessenger RNA half-life measurements in mammalian cells. *Methods Enzymol.* **448**, 335–357 (2008).
44. Bustamante, J., Boisson-Dupuis, S., Abel, L. & Casanova, J. L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin. Immunol.* **26**, 454–470 (2014).
45. Kumar, K. G. et al. Identification of adropin as a secreted factor linking dietary macronutrient intake with energy homeostasis and lipid metabolism. *Cell Metab.* **8**, 468–481 (2008).
46. Brigelius-Flohe, R. & Maiorino, M. Glutathione peroxidases. *Biochim. Biophys. Acta* **1830**, 3289–3303 (2013).
47. Kaur, S. J., McKeown, S. R. & Rashid, S. Mutant SOD1 mediated pathogenesis of amyotrophic lateral sclerosis. *Gene* **577**, 109–118 (2016).
48. Maiorino, M., Conrad, M. & Ursini, F. GPx4, lipid peroxidation, and cell death: discoveries, rediscoveries, and open issues. *Antioxid. Redox. Signal* **29**, 61–74 (2018).
49. Liang, H. et al. Short form glutathione peroxidase 4 is the essential isoform required for survival and somatic mitochondrial functions. *J. Biol. Chem.* **284**, 30836–30844 (2009).
50. Kawamata, H. & Manfredi, G. Import, maturation, and function of SOD1 and its copper chaperone CCS in the mitochondrial intermembrane space. *Antioxid. Redox Signal* **13**, 1375–1384 (2010).

Acknowledgements

We acknowledge funding from the National Key R&D Program of China (Ministry of Science and Technology, grant nos. 2017YFA0503600 and 2018YFA0507600), the National Natural Science Foundation of China (grant nos. 91753131 and 21675098), Beijing Natural Science Foundation (grant no. 5182011), the Interdisciplinary Medicine Seed Fund of Peking University (grant no. BMU2017MC006), Li Ge-Zhao Ning Life Science Junior Research Fellowship and Beijing Advanced Innovation Center for Structural Biology. P.Z. and J.W. are sponsored by the National Thousand Young Talents Award. We thank X. Chen (Peking University, Beijing, China), K. Sarkisyan (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia) and A.Y. Ting (Stanford University, USA) for providing plasmids, and J. Yang (Peking University, Beijing, China) for providing HEK293T cells. T. Ding assisted with photo-oxidation product characterization. C. Xiao assisted with streptavidin dot blot analysis. We thank Y. Wang, M. Conrad and Y. Sun for advice, and the National Center for Protein Sciences at Peking University in Beijing, China, for assistance with Fragment Analyzer 12.

Author contributions

P.W., W.T. and P.Z. conceived the project. P.W., W.T., Z.L., J.W. and P.Z. designed the experiments. P.W. and W.T. performed all experiments, unless otherwise noted. Z.L. and J.W. designed sequencing data analysis. Z.Z. and R.L. carried out mass-spec experiments. Y.Z. carried out probe synthesis. P.W. and T.X. carried out FISH experiments. P.W., W.T., Z.L., J.W. and P.Z. analyzed data. P.W., W.T. and P.Z. wrote the paper with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-019-0368-5.

Correspondence and requests for materials should be addressed to J.W. or P.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Reagents. Supplementary Table 1 summarizes the reagents used in this study. For chemical synthesis, D-biotin was purchased from TCI. Tert-butyl (2-aminoethyl) carbamate was purchased from Aladdin. (2-aminoethyl) Aniline, *N*-hydroxysuccinimide, *N,N′*-dicyclohexylcarbodiimide, 1-ethy-l-3-(3-dimethylaminopropyl) carbodi-imide hydrochloride, trimethylamine and *N*-boc-aminooxyacetic acid were purchased from J&K.

Plasmid construction. Supplementary Table 2 lists the genetic constructs used in this study. The miniSOG sequence was cloned from a gift plasmid from X. Chen (Peking University). The *MAVS* sequence was cloned from plasmid pcDNA3-Flag-APEX2-MAVS, a gift from A. Ting (Stanford University). Mitochondrial matrix targeting sequence was derived from the N-terminal 24-amino acid targeting sequence of *COX4*. Human *SEC61B* complementary DNA was obtained from the Harvard hORFeome Database. Genes were amplified via PCR and cloned into specified vectors by homemade Gibson assembly reagents. For bacterial expression, miniSOG was cloned into pET21a vector with 6× His-tag at the C-terminus. For mammalian expression, miniSOG fusion constructs were assembled into a pcDNA3.1(−) vector. To create cell lines stably expressing miniSOG fusion proteins, these miniSOG fusion genes constructs were subcloned into a lentiviral vector pLX304 (a gift from A. Ting). A V5 epitope tag (GKPIPNPLLGLDST) was added to these fusion constructs.

Recombinant protein expression and purification for in vitro CAP-seq labeling assay. *Escherichia coli* BL21(DE3) cells were transformed with pET21a-miniSOG. Bacteria were grown in 400 ml LB medium supplemented with 100 μg ml^{-1} ampicillin at 37 °C. When optical density (OD_{600}) reached 0.5, 0.5 mM IPTG was added to introduce protein expression and bacteria continued to grow at 18 °C overnight. Cells were gathered by centrifugation (5,000 r.p.m. for 10 min). After re-suspended in binding buffer (50 mM Tris, 300 mM NaCl, pH 7.8), cells were lysed by ultrasonication for 20 min. Supernatant was collected after centrifugation at 11,000*g* for 1 h at 4 °C, and was subsequently incubated with 1.5 ml Ni-NTA Agarose beads (Qiagen, no. 30210). The slurry was mixed on a rotator for 1 h at 4 °C, before being loaded into a column. Excess liquid was removed by gravity flow. Ni-NTA beads were then washed with 20 ml wash buffer (50 mM Tris, 300 mM NaCl, pH 7.8, 10 mM imidazole) and recombinant protein was eluted from beads with elution buffer (50 mM Tris, 300 mM NaCl, pH 7.8, 200 mM imidazole). The purified protein was dialyzed against phosphate-buffered saline (PBS) overnight and concentrated by ultrafiltration (Amicon Ultra-4 centrifugal filter, EMD Millipore, UFC801096). The protein sample was aliquoted and stored at −80 °C.

Identification of nucleoside photo-oxidation products by LC–MS. Here, 3 mM nucleoside (adenosine, guanosine, cytidine and uridine) and 60 μM purified photosensitizer protein (miniSOG, miniSOG2, SOPP, KillerRed, KillerOrange, SuperNova) were mixed in a 10 μl reaction volume. A 10 W blue light-emitting diode (LED) was used for samples containing miniSOG, miniSOG2 and SOPP. A 10 W red LED was used for samples containing KillerRed, KillerOrange and SuperNova. After 20 min LED illumination (24 mW cm^{-2}) at room temperature, the reaction mixture was diluted to 350 μl with ddH_2O and filtered with 0.22 μm MCE membrane filter (Navigator, NMF04–2). The sample was then analyzed by ultra-performance liquid chromatography–mass spectrometry (UPLC–MS) on a Waters Auto Purification LC–MS system (3100 Mass Detector, 2545 Binary Gradient Module, 2767 Sample Manager and 2998 Photodiode Array Detector). The system was equipped with a Waters C18 Sun Fire separation column (5 μm, 150 × 4.6 mm^2). Photo-oxidation products of guanosine were detected in the positive ion mode: *m/z* 245.0 (Iz, $[M+H]^+$) and *m/z* 316.0 (Sp, $[M+H]^+$).

For tandem mass spectrometry detection, the 10 μl reaction system was diluted with 990 μl deionized water and filtered with 0.22 μm MCE membrane filter. Next, 20 μl of the filtered solution was injected into LC–MS/MS. The solution was detected by triple-quadrupole mass spectrometry (AB SCIEX QTRAP 5500) in the positive ion multiple reaction-monitoring mode after separation by UPLC on a C18 column. The following masses were detected for multiple products: *m/z* 316.0 (parent) to 184.0 (fragmented at glycosidic bond) for Sp, *m/z* 263.1 to 131.1 for Oz, *m/z* 300.1 to 168.0 for 8-oxoG, *m/z* 245.0 to 113.0 for Iz and *m/z* 290 to 158.0 for Gh.

LC–MS/MS characterization of photo-oxidative conjugation of amine probe to guanosine. Here, 3 mM guanosine, 60 μM purified miniSOG and 5 mM amine-containing probe (Btn-NH_2 or PA) were mixed in 10 μl reaction volume. After illumination with blue LED for 40 min at room temperature, the reaction mixture was diluted to 1 ml with deionized water and then filtered with 0.22 μm MCE membrane filter. Then, 20 μl of filtered solution was injected into UPLC and detected by tandem mass spectrometry. The amine adducts of guanosine with biotin-conjugated alkylamine (Btn-NH_2) (*m/z* 672.0 to 540.1) and PA (*m/z* 353.0 to 221.0) was recorded.

Generation of cell lines stably expressing miniSOG fusions. HEK293T cells were cultured in DMEM medium (Invitrogen, C11995500BT) supplemented with 10% fetal bovine serum (Gibco, 10099044) and maintained at 37 °C with 5% CO_2. To prepare the lentivirus, HEK293T cells cultured in six-well plates were transfected at ~60% confluence with the gene of interest in lentiviral vector pLX304 (1 μg), together with two packaging plasmids, dR8.91 (1 μg) and pVSV-G (700 ng), and 10 μl PEI for 4 h. Then, 48 h after transfection, the culture medium containing lentivirus was collected and filtered through a 0.45 μm filter. After that, 1 m of the lentivirus-containing medium was added to a fresh HEK293T cells at ~70% confluence with ~600,000 cells. Next, 48 h after lentivirus transfection, the culture medium was exchanged to fresh complete medium containing 5 μg ml^{-1} blasticidin (Selleck, S7419) for selection. Infected cells were maintained in 5 μg ml^{-1} blasticidin-containing culture medium for 7 d, with fresh medium exchange every day. miniSOG expression in selected cells were verified via immunofluorescence. These cell lines were maintained in 5 μg ml^{-1} blasticidin-containing culture medium.

Fluorescence imaging. Cells were seeded on glass coverslips in 24-well plates at a density of ~70,000 cells per well. To improve the adherence of HEK293T cells, glass coverslips were pretreated with 20% Corning Matrigel matrix (Corning, no. 356234) diluted in DMEM for 20 min at 37 °C and washed with PBS (Solarbio, pH 7.2–7.4) once before use. After 24 h, cells were washed with PBS once, incubated with 10 mM PA in fresh Hanks Balanced Salt Solution (HBSS, Gibco, no. 14025092) for 5 min at 37 °C and then illuminated with blue LED for 20 min at room temperature. Thereafter, cells were washed with PBS once and fixed with 4% formaldehyde in PBS at room temperature for 15 min. Excess formaldehyde were removed from fixed cells through washing with PBS three times. Cells were then permeabilized with 0.2% Tween-20 (Solarbio, T8200) in PBS and then washed three more times with PBS. Next, 150 μl mixture of click reaction reagents was added to each well, containing 50 μM N_3-PEG_3-biotin (10 mM stock in DMSO), 2 mM CuSO4, 1 mM BTTAA and 0.5 mg ml^{-1} sodium ascorbate and incubated at room temperature for 30 min. After the click reaction, cells were washed with PBS three times and then blocked with 3% BSA in PBS (blocking buffer) for 30 min at room temperature.

For immunostaining, cells were incubated with primary antibody (mouse anti-V5 antibody at 1:1,000 dilution, rabbit anti-Hsp60 at 1:50 dilution or rabbit anti-Calnexin at 1:300 dilution) for 1 h at room temperature. After washed with PBST (0.2% Tween-20 in PBS) three times, cells were incubated with secondary antibody (goat anti-mouse Alexa Fluor 488 at 1:1,000 dilution, or goat anti-rabbit Alexa Fluor 568 at 1:1,000 dilution) and/or Streptavidin-Alexa Fluor 637 (ThermoFisher, S21374, 1:1,000 dilution) for 1 h at room temperature. Cells were then washed three times with PBST and counterstained with DAPI (ThermoFisher, D1306) in PBS for 5 min at room temperature. Cells were maintained in PBS for imaging after washed three times with PBS. Antibodies used in this study can be found in Supplementary Table 1.

Immunofluorescence images were collected with an inverted fluorescence microscope (Nikon-TiE) equipped with a spinning disk confocal unit (Yokogawa CSU-X1) and a scientific complementary metal-oxide semiconductor camera (Hamamatsu ORCA-Flash 4.0 v.2). The imaging equipment was controlled with a customized software written in LabVIEW v.15.0 (National Instruments).

Dot blot analysis of in vitro RNA labeling. Total RNA was extracted from HEK293T cells by TRIzol reagent (Invitrogen, 15596018) following the manufacturer's instructions. Then 200 ng μl^{-1} RNA was incubated with 100 μM photosensitizer (miniSOG, miniSOG2, SOPP, riboflavin or EosinY) and 5 mM amine-containing probe (biotin-conjugated probes or alkyne-conjugated probes), with or without 100 mM sodium azide in a 50 μl reaction in 200 μl PCR tube. After illumination with 10 W blue LED for 15 min at room temperature, RNA was purified from the reaction mixture with RNA Clean & Concentrator kit (Zymo, R1018). For experiments using alkyne probes, purified RNA was biotinylated via copper-assisted alkyne-azide cycloaddition reaction in the presence of 0.1 mM biotin-conjugated azide, 2 mM THPTA, 1 mM $CuSO_4$ and 1 mg ml^{-1} sodium ascorbate. The copper-assisted alkyne-azide cycloaddition reaction was incubated on a shaker at room temperature for 30 min. Thereafter, RNA was purified from the reaction mixture with RNA Clean & Concentrator kit and eluted into nuclease-free water.

Equal volume of purified biotinylated RNA was loaded onto Immobilon-Ny + membrane (Merck Millipore, INYC00010-1) and crosslinked to the membrane by an ultraviolet crosslinker (Analytik Jena). The membrane was blocked in 5% BSA at room temperature for 1 h and incubated with Streptavidin-HRP (Pierce, 21124) at room temperature for 1 h. The membrane was washed three times with TBST (50 mM Tris, 150 mM NaCl, 0.1% Tween-20, pH 7.4–7.6) for 10 min each time, incubated in Clarity Western ECL Substrate (Bio-Rad, 1705061) and then imaged on a ChemiDoc imaging system (Bio-Rad).

For Supplementary Fig. 7, the biotinylation yield was calculated as follows: the biotinylation intensity for 116 pmol purified oligo fell between the 1 and 5 pmol biotin-oligo standards. Thus, for every 23 (=116/5) to 116 (=116/1) oligoes, one biotin occurred in the in vitro miniSOG labeled oligoes. Considering the length of oligo was 28 nucleotides, the in vitro labeling yield was approximately once per 640 nt (23 × 28 nt) to once per 3,200 nt (116 × 28 nt). For Supplementary Fig. 13, 15 μg miniSOG-NES labeled total RNA showed a similar signal intensity to the 0.05 pmol biotin-oligo. Taking the average molecular weight of RNA nucleotides as 321, the labeling yield was calculated as $0.05/(15\times10^6/321) = 1/9.3\times10^5$ or approximately one biotin in 9.3×10^5 nucleobases.

RT–qPCR analysis of in vitro labeled RNA. Here, 2 μg in vitro labeled RNA was reverse transcribed with random primers, oligo dT and ProtoScript II (NEB, E6560L) in a 40 μl reaction according to the manufacturer's instructions. After cDNA synthesis, 5 μl of the reaction was taken as INPUT, while the remaining 35 μl was used for enrichment. Next, 40 μl Dynabeads MyOne Streptavidin C1 (Invitrogen, 65002) were washed three times with 200 μl loading buffer (100 mM Tris, pH 7.5, 1 M NaCl, 10 mM EDTA, 0.2% v/v Tween-20), re-suspended in 200 μl blocking buffer (1 μg μl^{-1} BSA and 1 μg μl^{-1} yeast tRNA in nuclease-free water) and incubated on a rotating mixer (ThermoFisher) at room temperature for 2 h. The blocked beads were washed once with loading buffer and re-suspended in 200 μl loading buffer. Then 35 μl of cDNA/RNA sample was added to blocked beads and the mixture was incubated at room temperature for 45 min on a shaker at 1,000 r.p.m. After removing supernatant, the beads were washed three times with wash buffer (100 mM Tris, pH 7.5, 4 M NaCl, 10 mM EDTA, 0.2% v/v Tween-20) at room temperature, twice with wash buffer at 50 °C for 10 min and twice with PBS. The beads were then re-suspended with 50 μl elution buffer containing 0.1 U μl^{-1} RNase H (NEB, M0297L), RNase A/T1 mix (Thermo Scientific, EN0551, at 40 ng μl^{-1} and 0.1 U μl^{-1}, respectively) and 12.5 mM D-biotin (Invitrogen, B20656). The mixture was incubated at 37 °C for 30 min on a shaker at 1,000 r.p.m. The supernatant containing eluted cDNA was collected and purified by DNA Clean & Concentrator kit (Zymo, D4013). The cDNA was eluted into 35 μl nuclease-free water (ENRICH).

For qPCR analysis, 0.5 μl INPUT or 4 μl ENRICH cDNA was added into each tube as the template. The templates were mixed with PowerUp SYBR Green Master Mix (Life, A25742) and quantified by ABI StepOne Plus system. C_t values were averaged from four replicate measurements. Negative controls with probes omitted were treated in the same manner as the sample, and were used here to calculate enrichment FC: $2^{\Delta Ct\_control - \Delta Ct\_label}$, where $\Delta C_t = C_{t\_ENRICH} - C_{t\_INPUT}$.

Labeling and purification of cellular RNA for RT–qPCR and next-generation sequencing analysis. Cells stably expressing miniSOG fusion genes were seeded to 100-mm cell culture dishes, with ~10^6 cells per dish. After 48 h, at ~90% confluence, cells were washed once with HBSS (Gibco, 14025092), incubated with 10 mM PA (Accela, SY002930) in HBSS at 37 °C for 10 min and illuminated with a 20 W blue LED for 20 min at room temperature. The sample was placed approximately 6–7 cm from the LED light source with an illumination intensity of 24 mW cm^{-2}, as measured at the sample plane. In this configuration, heating was negligible. Thereafter, cells were washed with PBS once and lysed immediately with TRIzol reagent. Briefly, the homogenized sample was mixed and incubated with chloroform, the upper aqueous phase was pipetted out and subjected to RNA precipitation by adding 100% isopropanol. The RNA pellet was washed with 75% ethanol and dissolved in RNase-free water. The purified total RNA was analyzed on Fragment Analyzer (Agilent) to detect the RNA integrity. Only RNA samples with RQN > 7 were subjected to downstream application.

For mitochondrial matrix labeling, total RNA was fragmented immediately without isolating mRNA. For ERM and OMM CAP-seq experiments, 600–1,000 μg total RNA was used to isolate mRNA with the Oligo $(dT)_{25}$ cellulose beads (NEB, S1408S) according to the manufacturer's instructions. Briefly, total RNA was denatured at 65 °C for 5 min, cooled in an ice bath and added to the Oligo $(dT)_{25}$ cellulose beads. After two rounds of poly-A mRNA binding, the mRNA-bound cellulose beads were washed and eluted in 20 mM Tris-HCl (pH 7.5).

The eluted mRNA was purified with RNA Clean & Concentrator kit (Zymo, R1018, size limits were 17 nt to 23 kilobases), and then subjected to fragmentation with NEBNext Magnesium RNA Fragmentation Module (RNA was fragmented at 94 °C for 5 min at the presence of magnesium). After purification again with RNA Clean & Concentrator kit, the fragmented RNA was digested by DNase I (NEB, M0303S) at 37 °C for 30 min. The mixture was incubated with 0.1 mM biotin-azide, 2 mM THPTA, 0.5 mM $CuSO_4$ and 5 mM sodium ascorbate on a shaker at room temperature for 10 min. After click reaction, RNA was purified again with RNA Clean & Concentrator kit and was eluted into nuclease-free water. Then, 200 ng purified mRNA (or 2 μg purified total RNA) was taken out and set aside as INPUT. The remaining RNA was purified with streptavidin-coated beads, as follows.

Mitochondrial drug perturbation. For CHX treatment, miniSOG-MAVS cells were incubated with HBSS containing 100 μg ml^{-1} CHX and 10 mM PA at 37 °C under 5% CO_2 for 10–20 min, then the cells were illuminated with blue LED (24 mW cm^{-2}) for 20 min at room temperature. For FCCP experiments, cells were exposed to FCCP for a total of either 5 or 40 min. For shorter treatment, miniSOG-MAVS cells were first incubated with HBSS containing 10 μM FCCP at 37 °C under 5% CO_2 for 5 min and the medium was changed to HBSS containing 10 mM PA for another 10 min incubation. The cells were then illuminated with blue LED (24 mW cm^{-2}) for 20 min at room temperature. For longer treatment, miniSOG-MAVS cells were incubated with HBSS containing 10 μM FCCP at 37 °C under 5% CO_2 for 15 min, PA was added to a final concentration of 10 mM and the cells were further incubated for another 5 min. After that, the cells were illuminated with blue LED (24 mW cm^{-2}) for 20 min at room temperature. Negative controls omitting the PA probe were performed in parallel. Then the samples were processed in the same way as described above.

Enrichment of labeled cellular RNA. Here, 40 μl Dynabeads MyOne Streptavidin C1 (Invitrogen, 65002) were washed three times with 200 μl loading buffer (100 mM Tris, pH 7.5, 1 M NaCl, 10 mM EDTA, 0.2% v/v Tween-20), re-suspended in 200 μl blocking buffer (1 μg μl^{-1} BSA, 1 μg μl^{-1} yeast tRNA in nuclease-free water) and incubated on a rotating mixer at room temperature for 2 h. The blocked beads were washed once with loading buffer, and then mixed with purified RNA in 200 μl loading buffer. The mixture was incubated at room temperature for 45 min on a shaking incubator at 1,000 r.p.m. After removing the supernatant, the beads were washed with wash buffer (100 mM Tris, pH 7.5, 4 M NaCl, 10 mM EDTA, 0.2% v/v Tween-20) three times at room temperature and then twice at 50 °C (10 min each). Finally, to elute RNA, beads were washed twice with PBS, re-suspended with 50 μl elution buffer (95% v/v formamide, 10 mM EDTA, pH 8.0, 1.5 mM biotin) and incubated at 65 °C for 5 min. To improve elution, samples were further incubated at 90 °C for 5 min. Thereafter, the supernatant was collected and diluted into 1 ml TRIzol reagent. Eluted RNA was purified according to the instruction and dissolved in 20 μl nuclease-free water (ENRICH). During purification, 1 μl (20 μg) glycogen (Life, 10814010) was added to the aqueous phase before adding isopropanol to assist RNA precipitation.

RT–qPCR analysis of enriched RNA. For each sample, INPUT mRNA (or INPUT total RNA) equivalent to 1% of the mRNA used for enrichment and 2 μl ENRICH RNA were reverse transcribed with random primers and ProtoScript II (NEB, E6560L) in 20 μl reaction buffer, according to the manufacturer's instructions. Briefly, template RNA, random primer, ProtoScript II enzyme mix and reaction mix were mixed and incubated at 25 °C for 5 min, 42 °C for 1 h and 80 °C for 5 min.

The INPUT and ENRICH cDNAs were aliquoted into six tubes (for six genes) as templates for qPCR. The templates were mixed with PowerUp SYBR Green Master Mix (Life, A25742) and primers (Supplementary Table 3), and then quantified by ABI StepOne Plus system. C_t values were averaged from four replicate measurements. Negative controls with probes omitted were treated in the same manner as the sample, and were used here to calculate enrichment FC: $2^{\Delta Ct\_control - \Delta Ct\_label}$, where $\Delta C_t = C_{tENRICH} - C_{tINPUT}$.

Library construction for next-generation sequencing. NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB, E7770) was used for Next-Generation Sequencing library construction. Then, 100 ng INPUT RNA or 13 μl ENRICH RNA was reverse transcribed in the mixture with 4 μl of First Strand Synthesis Reaction Buffer (5×), 1 μl Random Primer and 2 μl of First Strand Synthesis Enzyme Mix, following incubation at 25 °C for 10 min, 42 °C for 25 min, 70 °C for 15 min. Next, 8 μl of Second Strand Synthesis Reaction Buffer (10×), 4 μl of Second Strand Synthesis Enzyme Mix and 48 μl water were added to the first strand reaction product for second strand cDNA synthesis and incubated for 1 h at 16 °C. Double-stranded cDNA was purified from the reaction mixture using VAHTS DNA Clean Beads (Vazyme, N411) by 1.8× beads binding, two rounds of 80% ethanol wash and 50 μl 0.1× TE buffer elution. Then the 50 μl double-stranded cDNA was mixed with 7 μl of End Prep Reaction Buffer, 3 μl Enzyme Mix and incubated at 20 °C for 30 min and at 65 °C for 30 min for end prep. For adaptor ligation, 2.5 μl Adaptor (linearized), 1 μl ligation Enhancer and 30 μl Ligation Master Mix were added to the 60 μl end prepped DNA and incubated for 20 min at 20 °C. The ligation reaction was purified using 0.9× VAHTS DNA Clean Beads. Adaptor ligated DNA was amplified by PCR, and purified again with 0.9 × VAHTS DNA Clean Beads. During PCR, 25 μl Q5 Master Mix, 2 μl primer mix and 23 μl adaptor ligated DNA were mixed and placed on a thermocycler with the following cycling conditions: 98 °C 30 s for one cycle, 98 °C 10 s with 65 °C 75 s for 11 (INPUT samples) or 15 (ENRICH samples) cycles, 65 °C 5 min for one cycle, held at 4 °C. The quality of purified libraries was assessed by Fragment Analyzer (Agilent). For the samples that showed a wide peak distribution, a size-selection step was performed using two-round VAHTS DNA Clean beads purification with 0.8× and 0.2× volume of beads. Every eight constructed libraries were equally mixed to 2 ng μl^{-1} and sent for high-throughput sequencing.

Next-generation sequencing and data analysis. All sequencing data analyzed during this study are included in the Source Data. All cDNA libraries were deep-sequenced for 150 base paired reads on the Illumina HiSeq X Ten platform. The sequencing reads for each library were mapped using HISAT2 (ref. [51]) (v.2.1.0) against the human genome assembly GRCh38 (hg38), which was downloaded from Ensembl. With the matching gene annotation (v.87) from Ensembl, the read counts of each gene were measured by HTSeq[52] (v.0.6.1), using the option '–stranded = no'. The raw counts of each gene were converted to FPKM values, which are summarized in Source Data Figs. 2–4 and 6a,d.

To analyze the mitochondrial matrix CAP-seq data, we calculated FCs of genes with input count > 100:

$$\log_2(\text{enrichment versus control}) = \log_2\left[\frac{\text{average FPKM}_{\text{label}}}{\text{average FPKM}_{\text{omit PA}}}\right]$$

and

$$\log_2(\text{enrichment versus input}) = \log_2\left[\frac{\text{average FPKM}_{\text{post-enrichment}}}{\text{average FPKM}_{\text{pre-enrichment}}}\right]$$

The genes with $\log_2$ (enrichment versus control) > 1 and $\log_2$ (pre- versus post-enrichment) > 2 were identified as the enriched in the mitochondrial matrix dataset (Source Data Fig. 2).

To analyze ERM and OMM CAP-seq data, differential expression analysis was carried out by DESeq2 (ref. [53]) (v.1.16.1). Raw counts were normalized to remove the effect of sequencing depth. FCs and *P* values for each gene across different treatments were reported. The cutoff for $\log_2$ (enrichment versus control) in the ERM CAP-seq dataset was determined using ROC analysis[54]. True positive RNAs were defined as those predicted by Phobius[54] to contain secretory signals. False positive RNAs were defined as those not predicted by Phobius[54], SignalP[55] and TMHMM[56] to contain any secretory signals. This definition is the same as that used in previous ribosome profiling study. To improve the quality of analysis, we have excluded genes at low expression levels (total INPUT counts fewer than 100 in replicate experiments) and genes with $\log_2$ (pre- versus post-enrichment) < 0.4. A cutoff of $\log_2$ (enrichment versus control) > 1.12 was determined to optimize precision (0.92) and accuracy (0.78). Entries with $P > 0.05$ were removed from the dataset. The final list of enriched genes is summarized in Source Data Fig. 3.

GOCC-secretome was defined as the ensemble of genes whose GOCC annotations include various subcellular locations related to the secretory pathway, including ER, Golgi apparatus, vesicles, plasma membrane and extracellular region. Supplementary Table 4 lists specific gene ontology terms used to define secretome in this study. Genes in the MitoCarta v.2.0 database were excluded from the secretome list. GOCC annotations of all human genes were downloaded from Gene Ontology Consortium (http://current.geneontology.org/products/pages/downloads.html).

HPA-secretome was generated from the HPA database. Genes with the following annotations were included in the list: 'Endoplasmic Reticulum', 'Golgi Apparatus', 'Plasma Membrane', 'Secreted proteins' and 'Vesicles'.

For basal OMM CAP-seq dataset (no drug treatment), the enrichment cutoff was set as $\log_2$ (enrichment versus control) > 0.5 and $P < 0.01$ (Source Data Fig. 4). For CHX-treated OMM CAP-seq dataset, the enrichment cutoff was set as $\log_2$ (enrichment versus control) > 0.5 and $P < 0.01$ (Source Data Fig. 6a). When compared with basal OMM data, all mRNA transcripts enriched in CHX-treated OMM CAP-seq dataset were considered and $\Delta\log_2$(FC) (CHX/Basal) of each enriched mRNA was calculated. mRNAs with $\Delta\log_2$(FC) (CHX/Basal) > 0.8 were defined as highly upregulated, mRNAs with $\Delta\log_2$(FC) (CHX/Basal) between 0.5 and 0.8 were defined as upregulated and those with $\Delta\log_2$(FC) (CHX/Basal) < 0.5 were defined as others. For the FCCP-treated OMM CAP-seq dataset, the enrichment cutoff was set as $\log_2$ (enrichment versus control) > 0 and $P < 0.05$ (Source Data Fig. 6d). All mRNA transcripts enriched in basal OMM CAP-seq dataset were used for comparison with FCCP-treated OMM data and $\Delta\log_2$(FC) (FCCP/Basal) was calculated. mRNAs with $\Delta\log_2$(FC) (FCCP/Basal) < −0.5 were defined as highly downregulated, those with $\Delta\log_2$(FC) (FCCP/Basal) > −0.5 were defined as others. MitoCarta2.0 was used for mitochondrial annotation. Sub-mitochondrial localization information was obtained from GOCC annotations (GO:0005759 for matrix, GO:0005743 for IMM, GO:0005758 for IMS and GO:0005741 for OMM). Genes were marked as 'multi' if multiple sub-mitochondrial localizations were annotated in GOCC. OXPHOS genes were also identified from GOCC annotations, including mitochondrial respiratory chain complexes (GO:0005747, GO:0005749, GO:0005750 and GO:0005751) and mitochondrial respirasome (GO:0005746). Ribosome subunits were annotated according to reported human 80S ribosome.

The sequencing coverage of each site on the interested genes was obtained by the SAMtools depth program, only counting the reads with a mapping quality greater than 60. Average coverages were calculated between all replicates and showed in the genome track: 'enrich' represents post-enrichment labeling samples, 'input' represents pre-enrichment labeling sample and 'control' represents post-enrichment sample with probe omitted.

To generate heat maps to compare RNA enrichment levels across experiments, the R package 'pheatmap' was used and the rows were scaled and clustered by the default method 'hclust'.

smiFISH verification of the OMM CAP-seq enriched RNAs. The sequences of smiFISH probes were designed on Oligostan[57] and summarized in Supplementary Table 5. Probes for each RNA transcript were prepared in equimolar mixture with 0.833 μM final concentration of individual probes. The secondary probe conjugated with Alexa Fluor 488 was dissolved in TE buffer at final concentration of 100 μM. The hybridization of 2 μl probe-set with 0.5 μl FLAP was performed in a PCR machine with 1 μl 10× NEB 3.1 buffer in 10 μl reaction following the procedure: 85 °C 3 min, 65 °C 3 min and 25 °C 5 min. HEK293T cells were plated on glass bottom coverslips in 24-well plates and cultured to reach 60–70% confluence. Mitotracker Deep Red 633 was added to the cell medium to a final concentration of 2 μM and incubated for 30 min at 37 °C. The cells were rinsed once with PBS and fixed with freshly prepared 4% paraformaldehyde in PBS for 20 min. After washing twice with PBS, the cells were permeabilized with 70% ethanol overnight at 4 °C, washed once again with PBS and then incubated in freshly prepared 15% formamide in 1× SSC buffer for 15 min. Mix1 (5 μl 20× SSC, 1.7 μl 20 μg μl^{-1} yeast tRNA, 15 μl formamide, 2 μl hybridized FLAP complex, 26.3 μl H_2O) and Mix2 (1 μl 20 mg ml^{-1} BSA, 1 μl 200 mM VRC, 26.5 μl 40% dextran sulfate, 21.5 μl H_2O) were added together and vortexed to prepare hybridization mix. Then, 50 μl of the hybridization mix was dropped on a 100 mm cell culture dish and the coverslip was laid on the drop with cells facing down. Following that, 1 ml 15% formamide/1× SSC buffer in a 35 mm dish was put inside the 100 mm dish and wrapped up with parafilm to prevent evaporation. Next, the 100 mm dish was incubated at 37 °C overnight. The coverslip was placed in a 24-well plate and washed twice with freshly prepared 15% formamide/1× SSC buffer for 30 min at 37 °C. Then the cells were rinsed twice with PBS and imaged on an inverted fluorescence microscope (Nikon Ti-E) with a ×60 oil-immersion objective.

To calculate the correlation coefficient of the colocalization of FISH signal with mitotracker, the original images with 1,024 × 1,024 pixels were loaded into ImageJ and analyzed by the 'Colocalization' function under 'Analyze' (check 'Colocalization Threshold'). Rcoloc values were used. Zoom-in views of 400 × 400 pixels were cropped to display in Fig. 5b.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data presented in this study are available in the main text and Supplementary Information. Sequencing data are included in Source Data Figs. 2–4 and 6a,d.

References

51. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
52. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Kall, L., Krogh, A. & Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).
55. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
56. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
57. Tsanov, N. et al. smiFISH and FISH—a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).