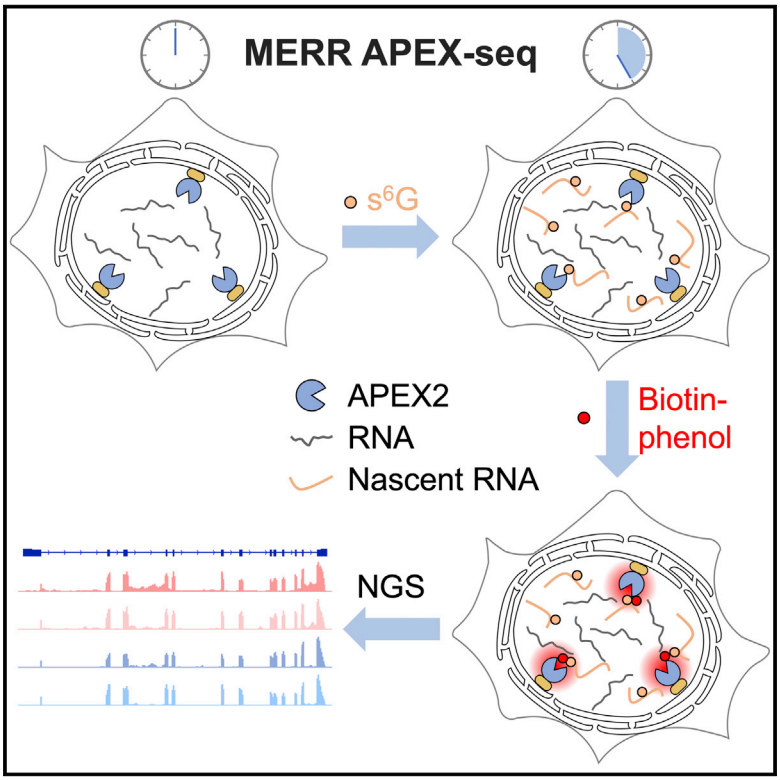
Resource

Cell Chemical Biology

Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome

Graphical abstract



Highlights

- Improving APEX-seq sensitivity via non-canonical ribonucleoside metabolic incorporation
- Selectively capturing newly transcribed RNA at subcellular locales in cultured cells
- Revealing nuclear lamina-associated mRNAs involved in controlling gene expression

Authors

Ran Li, Zhongyu Zou, Wentao Wang, Peng Zou

Correspondence

zoupeng@pku.edu.cn

In brief

Li et al. develop a sensitive method for the subcellular profiling of newly transcribed RNA in live cells, which uncovers nuclear lamina-associated mRNAs that encode for nuclear proteins involved in histone modification, chromosomal structure maintenance, and RNA splicing and processing.

Li et al., 2022, Cell Chemical Biology *29*, 1–14
July 21, 2022
https://doi.org/10.1016/j.chembiol.2022.02.005

CellPress

Cell Chemical Biology

CellPress

Resource

Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome

Ran Li,[1] Zhongyu Zou,[2,5] Wentao Wang,[2] and Peng Zou[1,2,3,4,6,*]

[1]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
[2]College of Chemistry and Molecular Engineering, Synthetic and Functional Biomolecules Center Beijing National Laboratory for Molecular Sciences, Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, Peking University, Beijing 100871, China
[3]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China
[4]Chinese Institute for Brain Research (CIBR), Beijing 102206, China
[5]Present address: Department of Chemistry, The University of Chicago, Chicago, IL, USA
[6]Lead contact
*Correspondence: zoupeng@pku.edu.cn
https://doi.org/10.1016/j.chembiol.2022.02.005

SUMMARY

The spatial arrangement of newly synthesized transcriptome in eukaryotic cells underlies various biological processes including cell proliferation and differentiation. In this study, we combine metabolic incorporation of electron-rich ribonucleosides (e.g., 6-thioguanosine and 4-thiouridine) with a peroxidase-mediated proximity-dependent RNA labeling technique (APEX-seq) to develop a sensitive method, termed MERR APEX-seq, for selectively profiling newly transcribed RNAs at specific subcellular locations in live cells. We demonstrate that MERR APEX-seq is 20-fold more efficient than APEX-seq and offers both high spatial specificity and high coverage in mitochondrial matrix. At the ER membrane, 91% of the transcripts captured by MERR APEX-seq encode for secretory pathway proteins, thus demonstrating the high spatial specificity of MERR APEX-seq in open subcellular compartments. Application of MERR APEX-seq to the nuclear lamina of human cells reveals a local transcriptome of 1,012 RNAs, many of which encode for nuclear proteins involved in histone modification, chromosomal structure maintenance, and RNA processing.

INTRODUCTION

The subcellular localization of newly transcribed RNAs has been widely observed in many cell types and plays a critical role in gene expression regulation, structural support, stress response, local protein translation, and development (Buxbaum et al., 2015; Khong et al., 2017; Yu et al., 2018). Newly synthesized RNA molecules undergo splicing and polyadenylation in the nucleus and navigate through the complex cellular milieu by interacting with an array of RNA-binding proteins (Kohler and Hurt, 2007; Mofatteh and Bullock, 2017). In metazoan cells, the dense fibrillar meshwork around inner nuclear membrane, called nuclear lamina (NL), is essential for organizing genomic structures and regulating RNA transcription (Tang et al., 2008). Genome regions contacting the NL, known as lamina-associated domains (LADs) (van Steensel and Belmont, 2017), have been found to form a repressive environment that features low gene expression (Akhtar et al., 2013; Kind et al., 2013). For example, lamin B receptor (LBR) interacts with the non-coding RNA *XIST* to recruit the X chromosome to the NL, which is crucial for inactivating the X chromosome (Chen et al., 2016; Engreitz et al., 2016). Mutations in NL-related genes have been linked to human diseases, such as laminopathies and autosomal dominant leukodystrophy (Dobrzynska et al., 2016; Padiath, 2019). Transcriptome-wide profiling of RNA at the NL could contribute to our understanding of the functional outcome of subcellular RNA localization.

To date, several methods have been developed to study the subcellular localization of RNA. Technologies based on fluorescence *in situ* hybridization (FISH) offer high spatial resolution but only allow analysis of a handful of genes at a time (Raj and Tyagi, 2010). Recent development of highly parallelized methods (e.g., multiplexed error-robust FISH [Chen et al., 2015]) has increased the analysis throughput and allowed the simultaneous examination of over 1,000 genes. However, all of these methods require prior knowledge of RNA sequences under investigation, and thus are unsuitable for bulk discovery-oriented RNA profiling (Tsanov et al., 2016). To gain understanding of RNA identities in specific subcellular regions, cellular fractionation methods have been routinely used to isolate the complete transcriptome. For example, physical separation of neuropil regions has allowed the investigation of RNA transport in neuronal processes (Kim

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

and Jung, 2015). Biochemical fractionations and fluorescence-activated particle sorting have been applied to isolations of membrane-bound (e.g., nucleus) and membrane-less (e.g., processing body) organelles, respectively (Hubstenberger et al., 2017; Lefebvre et al., 2017). However, these fractionation methods are often prone to contamination and are not generally applicable to structures that are difficult to purify (e.g., NL) (Wilkie and Schirmer, 2008).

More recently, proximity labeling techniques have been developed to label RNAs at specific subcellular locations. APEX2 is an engineered peroxidase that converts biotin-phenol (BP) substrate into a highly reactive phenoxyl free radical, which then reacts with nearby electron-rich biomolecules to form a covalent bond. While originally designed as a protein labeling tool, APEX2 has been later repurposed to label the guanosine nucleobase in RNA (APEX-seq), enabling the high-resolution mapping of subcellular transcriptome in live cells (Fazal et al., 2019; Padron et al., 2019). Alternatively, proximity-dependent RNA labeling could be achieved by miniSOG-mediated photo-oxidative conjugation of an affinity probe to guanosine in live cells (Wang et al., 2019). These methods have opened up new opportunities to study the spatial organization of RNA and have been successfully implemented in multiple subcellular organelles. However, the feasibilities of these methods are hampered by the large amount of input samples required owing to the low reactivity of APEX-generated phenoxyl free radical toward nucleobases or the ordinary ability of miniSOG to generate reactive oxygen species.

Given the electrophilic nature of phenoxyl free radicals, APEX is more suited for labeling electron-rich amino acid residues (e.g., tyrosine) in proteins rather than the electron-deficient nucleobases in RNAs. Indeed, APEX-seq primarily targets guanosine, which has the lowest redox potential among all four ribonucleobases (Fazal et al., 2019). Yet the efficiency of APEX-mediated RNA labeling is still quite low, occurring approximately once every 50,000 nucleosides (Zhou et al., 2019). Efforts to screen alternative APEX substrates identified biotin-aniline as a more efficient probe for labeling RNA (Zhou et al., 2019). The strategy of metabolically incorporating 4-thiouridine to enhance RNA biotinylation with the assistance of APEX2 was applied in mitochondrial matrix to detect the transcripts restricted in mitochondrial membrane, but there were no further studies employed in open regions (Huang et al., 2020). In this study, we sought to improve the sensitivity of APEX-seq by metabolically incorporating electron-rich ribonucleosides into newly transcribed RNA molecules (MERR APEX-seq). Following the metabolic incorporation of 6-thioguanosine (s^6G) or 4-thiouridine (s^4U) into transcripts, APEX-mediated proximity-dependent RNA biotinylation is triggered by hydrogen peroxide, and the biotinylated transcripts are subsequently isolated and sequenced. MERR APEX-seq can increase the level of enrichment by 20-fold and can be applied to various subcellular compartments including the mitochondrial matrix, the ER membrane, and the NL. In the mitochondrial matrix, MERR APEX-seq identifies all 13 mitochondrial (mt)-mRNAs and both mt-rRNAs. At the ER membrane (ERM), 91% of the transcripts captured by MERR APEX-seq encode for secretory pathway proteins, thus demonstrating the high spatial specificity of MERR APEX-seq in open subcellular compartments. At the NL, MERR APEX-seq unveils hundreds of transcripts that were not previously covered in the APEX-seq dataset, with a majority of them encoding for nuclear proteins involved in histone modification, chromosomal structure maintenance, and RNA splicing and processing.

RESULTS

Electron-rich ribonucleosides facilitate APEX-mediated RNA labeling

Non-canonical ribonucleosides could be metabolically incorporated into the cellular transcriptome to provide bio-orthogonal functional handles for the selective labeling of biomolecules. For example, photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation employs s^4U to enhance the efficiency of RNA-protein photo-crosslinking (Hafner et al., 2010; Melvin et al., 1978). Metabolic RNA labeling is also useful for specifying nascent transcriptome, as exemplified in TimeLapse-seq, where s^4U-tagged nascent RNAs are identified through chemical oxidation and mutation signatures in sequencing (Kiefer et al., 2018; Schofield et al., 2018). In this study, we chose s^6G and s^4U because these are substantially more electron rich than the natural ribonucleosides, as indicated by their lower redox potential (0.55 V and 0.63 V for s^6G and s^4U, respectively, compared with 1.29 V for G) (Arias et al., 2006; Heihoff et al., 1990; Holzer and Wrona, 1983; Steenken and Jovanovic, 1997).

We started by measuring the metabolic incorporation rate of s^6G and s^4U in cultured human embryonic kidney 293T (HEK293T) cells. Liquid chromatography-tandem mass spectrometry analysis revealed 0.52% substitution rate for s^6G in place of guanosine in the total RNA extract after cells were incubated with 100 μM s^6G in the culturing medium for 5 h (Figure S1A and Table S1). The metabolic incorporation rate of s^4U was slightly higher, with 1.17% in total RNA lysate following incubation with 100 μM s^4U for 5 h (Figure S1B and Table S2). We went on to test whether s^6G/s^4U incorporation could lead to increased APEX-BP labeling efficiency. We incubated human embryonic kidney 293T (HEK293T) cells with 100 μM non-canonical ribonucleoside in the culturing medium for 5 h (Figure 1A) and performed APEX-mediated BP labeling on extracted total cellular RNA *in vitro*. As a negative control, APEX2, BP probe, or H_2O_2 were omitted from the reaction mixture. Streptavidin-horseradish peroxidase (HRP) dot blot analysis revealed that biotinylation efficiency was 9- and 3-fold higher for s^6G- or s^4U-containing RNAs, respectively, than RNAs without metabolic incorporation (Figures 1B and S1C). APEX-mediated BP labeling of s^6G- or s^4U-containing RNAs is dependent on APEX2, BP probe, and H_2O_2. Quantitative dot blot analysis with synthetic s^4U-containing oligonucleotide revealed that BP labeling occurs approximately once every 300 s^4U (Figures S1D–S1G). Therefore, we validated electron-rich nucleoside incorporation as an effective approach to enhance APEX-mediated RNA biotinylation *in vitro*.

We then evaluated the sensitivity of MERR APEX-seq in the context of live cells. We chose the mitochondrial matrix as a model because of its well-defined transcriptome, which included 13 mt-mRNAs, 2 mt-rRNAs, and 22 mt-tRNAs (Mercer et al., 2011). We constructed a HEK293T cell line stably expressing APEX2 in the mitochondrial matrix (MITO-APEX2). s^6G or s^4U

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

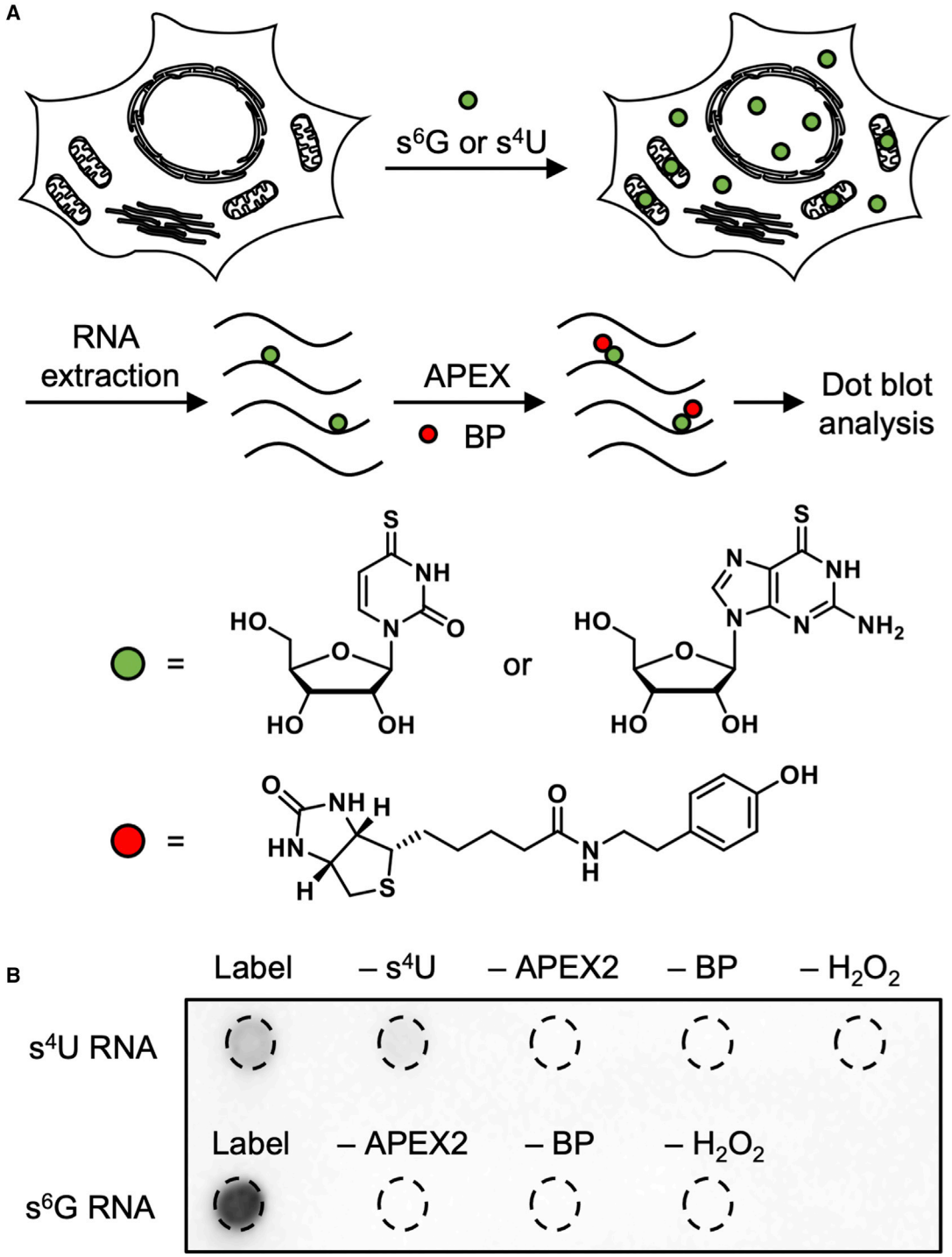


Figure 1. Development of APEX-mediated electron-rich ribonucleoside labeling strategy *in vitro*

(A) Scheme of APEX-mediated metabolic RNA labeling *in vitro*. Following s^6G/s^4U metabolic incorporation, cellular RNAs were extracted and biotinylated by APEX2 in the presence of 0.5 mM BP and 1 mM H_2O_2 for 1 min. Biotinylated RNAs were purified and analyzed by streptavidin-HRP blotting.

(B) Enhanced APEX-mediated RNA biotinylation by incorporation of electron-rich nucleosides. Biotinylation level was measured by streptavidin-HRP blotting. The figure shown is representative of two independent experiments.

See also Figure S1; Tables S1 and S2.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

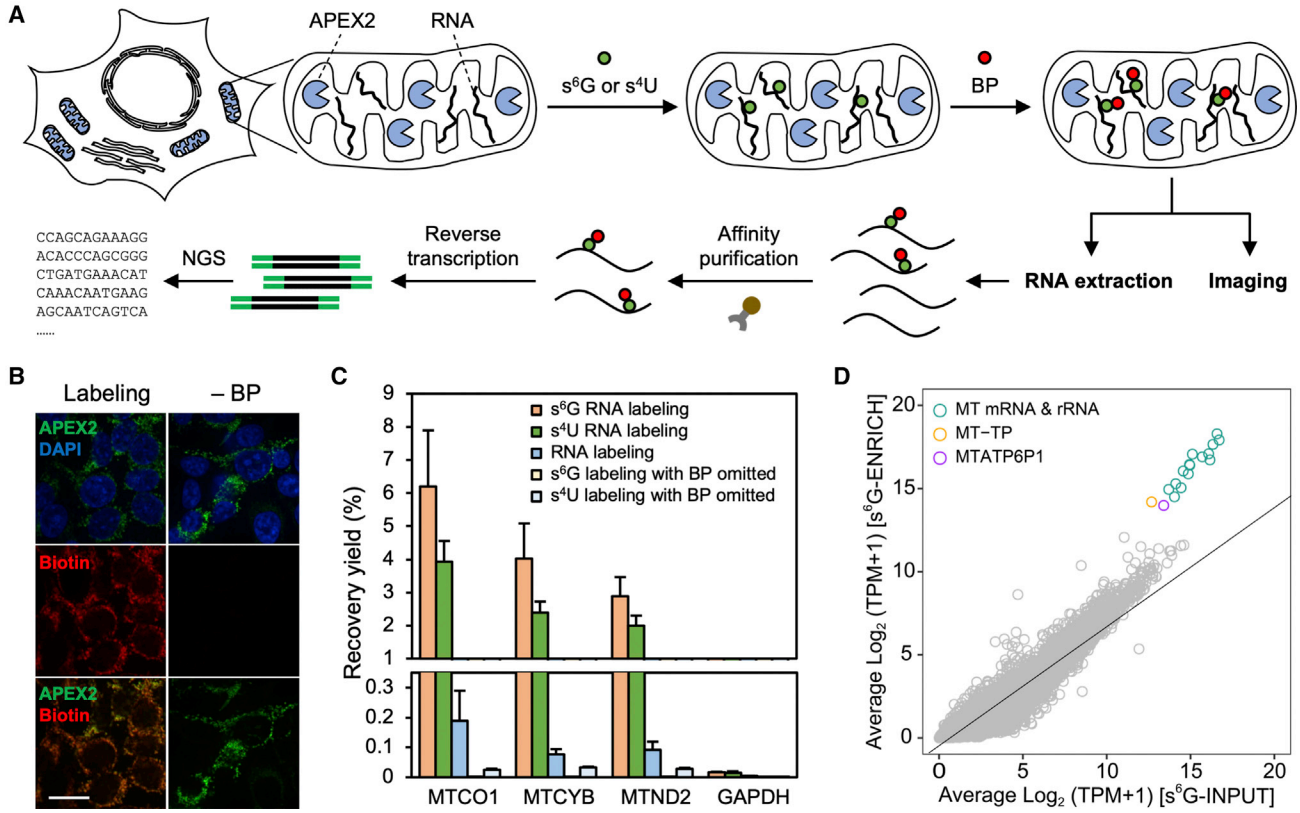


Figure 2. Characterization of MERR APEX-seq in mitochondrial matrix

(A) Scheme of MERR APEX-seq workflow. APEX2 is genetically targeted to the mitochondrial matrix (MITO) through N-terminal fusion with the mitochondrial targeting sequence derived from human COX4. Cultured cells are incubated with s^6G/s^4U and BP probe before the labeling reaction is triggered by the addition of 1 mM H_2O_2 for 1 min. Labeled cells are either fixed for imaging analysis or lysed for enrichment, followed by qRT-PCR analysis and next-generation sequencing.

(B) Representative immunofluorescence images of MITO-APEX2 labeling in HEK293T cells. Cells are incubated with s^6G and labeling reagents as described in (A). Negative controls omitting the BP probe are also shown for comparison. Fluorescence intensities in APEX2 and SA-647 channels have been normalized across different experimental conditions. These results have been repeated twice independently. Scale bar, 20 μm.

(C) qRT-PCR analysis of the enrichment yields for MT-mRNAs. Recovery rate is calculated from the C_t values of ENRICH and INPUT samples across four technical replicates.

(D) Scatterplot analysis comparing the ENRICH versus the INPUT samples of s^6G MERR APEX-seq. MT-mRNAs and MT-rRNAs are represented as 15 green dots. The orange dot represents an mt-tRNA, *MT-TP*, and the purple dot is a mitochondrial pseudogene. The black line indicates the least-square fit of all data points.

See also Figure S2; Data S1 and S6.

was supplied in the culturing medium for 5 h (Hafner et al., 2010), followed by 30 min of BP incubation and pulse treatment with hydrogen peroxide for 1 min. The labeling was immediately quenched by a cocktail of free radical scavenger (10 mM sodium ascorbate) and peroxidase inhibitor (5 mM Trolox and 10 mM sodium azide). Thereafter, cells were either fixed for immunofluorescence analysis or lysed for RNA extraction (Figure 2A). Immunofluorescence imaging revealed that the biotinylation signal colocalizes with APEX2 expression (Figures 2B, S2A, and S2B). In negative controls omitting APEX2 expression or BP probe, negligible background was observed. Following cell lysis, extracted total RNA was digested by DNase I to remove residual contaminating DNA (INPUT samples) and quality controlled with Fragment Analyzer. Only samples with an RNA quality number (RQN) >8.0 were used for subsequent analysis. Biotinylated RNA was enriched by streptavidin-coated magnetic beads (ENRICH samples). Both the INPUT and immunoprecipitated RNAs were reverse transcribed with SuperScript III and analyzed by qRT-PCR. Consistent with the observed higher labeling efficiencies *in vitro*, cellular samples that underwent s^6G or s^4U metabolic incorporation exhibited 39- or 25-fold higher levels of mt-mRNA enrichment, respectively, as compared with untreated samples. In all experiments, the enrichment yield of the cytosolic RNA marker *GAPDH* was at least 132-fold lower than those of mt-mRNAs, demonstrating the high spatial specificity of our labeling strategy (Figures 2C and S2C).

Aiming to optimize the labeling method also from the biotin-containing probe side, we made use of our previous finding that APEX-mediated RNA labeling is more efficient with biotin-aniline (BA) probe than with BP probe (Zhou et al., 2019) and compared their MERR APEX labeling efficiencies. For both biotin probes, dot blot analysis shows a substantially higher level of biotinylation in RNA samples containing metabolically incorporated s^4U/s^6G than the control samples omitting the metabolic

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

labeling *in vitro*. Regardless of whether containing s^4U/s^6G or not, samples labeled with BA are overall more strongly biotinylated than samples labeled with BP (Figure S2D). Consistent with the above characterization *in vitro*, qRT-PCR analysis of cellular labeling in HEK293T cells expressing MITO-APEX2 reveals that, for both BA and BP probes, the recovery yields of mt-mRNA markers (*MTCO2*, *MTCYB*, *MTND2*) relative to the cytosol RNA marker (*GAPDH*) are substantially improved following s^6G/s^4U incorporation (Figure S2E). Interestingly, MERR APEX labeling with BP achieves higher recovery yields than labeling with BA. This discrepancy between *in vitro* and cellular labeling may arise from the presence of free radical quenchers (e.g., thiols, phenols) in the native cellular environment. Overall, we concluded from the above characterizations that MERR APEX labeling with BP is more efficient, a corollary we applied for all subsequent experiments.

After validating the sensitivity and specificity of MERR APEX-seq, we constructed cDNA libraries with RNAs enriched in the mitochondrial matrix (Data S1). Duplicated MERR APEX-seq experiments were performed, and the transcripts per kilobase million (TPM) values of enriched RNAs were highly correlated across replicated experiments (Pearson's correlation coefficient >0.95) (Figures 2A and S2F–S2H). We further confirmed that the metabolic incorporation of s^6G did not disturb gene expression levels (Figures S2I–S2L). As shown in Figure 2D, a scatterplot showing the averaged TPM values in the ENRICH versus the INPUT reveals the enrichment of 17 RNAs in MERR APEX-seq, including all 13 mt-mRNAs, both mt-rRNAs, one mt-tRNA *MT-TP*, and a pseudogene *MTATP6P1*, all of which were transcribed from the mitochondrial genome. Finally, when cells were incubated with 100 μM s^4U for 5 h, MERR APEX-seq yielded a similar pattern of enrichment (Figures S2M and S2N). To investigate whether MERR APEX labeling could cause nucleobase mutations, we analyzed the occurrence of SNPs in our mitochondrial labeling datasets. While G-to-A mutation was rarely observed in s^6G MERR APEX-seq, we did notice T-to-C mutation in s^4U MERR APEX-seq, which likely arose from s^4U oxidation (Table S3). The mutation rate was independent of BP labeling. Given its higher labeling efficiency toward BP and lower mutation frequency, we chose s^6G for further MERR APEX-seq experiments. Collectively, our data established MERR APEX-seq as a next-generation sequencing (NGS)-compatible method to profile mitochondria-associated transcriptome with high sensitivity and specificity. The wide use of APEX2-mediated proteome profiling indicates that MERR APEX-seq bears the potential to be employed in the study of other cellular organelles.

Enrichment of transcriptome at ER membrane demonstrates high spatial specificity of MERR APEX-seq

To evaluate the spatial specificity of MERR APEX-seq more rigorously, we chose the ERM as an example for open subcellular space. As most secretory pathway proteins (proteins targeted to the ER, Golgi apparatus, the plasma membrane, and secreted proteins) are initially synthesized at the ERM, it is expected that ribosome-bound mRNAs encoding for these proteins would be enriched. Thus, the ERM is a convenient benchmark for characterizing the specificity and sensitivity of RNA proximity labeling techniques (Kaewsapsak et al., 2017). We targeted APEX2 to the ERM in HEK293T cells via fusion with the β subunit of ER translocon complex Sec61 (APEX2-ERM) (Figures 3A and 3B). As a non-targeted control, we localized APEX2 to the cytoplasm by fusing it with a nuclear export sequence (APEX2-NES) derived from residues 6–17 of the HIV-1 Rev protein 30 (Figures 3C and 3D).

We performed four replicated MERR APEX-seq experiments at the ERM and two at the NES. Pearson's correlation coefficients of these experiments showed high similarities among ENRICH samples in both subcellular regions (Figure S3A). We applied receiver-operating characteristic curve (ROC) analysis to determine the threshold of the $\log_2$ fold change (ENRICH versus INPUT) to be 0.27 (Figure S3B; Data S2 and S3). We further generated a high-confidence dataset with p value less than 0.05 and filtered out transcripts with low abundance by removing those with baseMean values under 100. Together, the above cutoffs yielded a list of 1,035 transcripts, including 1,011 mRNAs (98%) (Figures 3E, S3C, and S3D; Data S2). Gene ontology (GO) analysis revealed that these mRNAs predominantly encoded for proteins localized in the secretory pathway (Figure S3E). To evaluate the specificity of the ERM MERR APEX-seq dataset, we compiled a list of secretome genes using combined knowledge from Gene Ontology Cellular Component (GOCC) (Ashburner et al., 2000; Carbon et al., 2009; Day-Richter et al., 2007; Mi et al., 2019; The Gene Ontology, 2019) and the Human Protein Atlas (Uhlen et al., 2015), and found that 91% (943 out of 1,035) had secretome annotations, similar to previously reported specificity of APEX-seq at the ERM (89%) (Figure 3F). The above analysis shows that MERR APEX-seq offers exceptional spatial specificity at open subcellular regions.

Notably, our MERR APEX-seq dataset at the ERM showed enrichment of a distinct pool of transcripts as compared with a published APEX-seq dataset (Fazal et al., 2019) at the same subcellular location, indicating a difference in the bias of the two methods (Figure S3F). It is expected that MERR APEX-seq with s^6G preferentially targets transcripts with higher content of guanosine. Consistent with this expectation, the GC content of the ERM MERR APEX-seq dataset is significantly higher than that of the APEX-seq dataset (48.2% versus 43.4%, Figure S3G). Besides, no statistically significant differences are found between the length distribution of transcripts enriched in the ERM MERR APEX-seq dataset versus that of secretory genes (Figure S3H). Analysis of transcript abundance in ERM MERR APEX-seq and ERM APEX-seq datasets (using baseMean values of DESeq2 as a proxy) reveals that transcripts captured by both methods overall have higher baseMean values than those captured by only one method, indicating that lowly expressed transcripts captured by MERR APEX-seq and APEX-seq are different (Figure S3I).

Profiling nascent transcriptome at the nuclear lamina reveals chromatin-related functional features

It is estimated that more than one-third of the mammalian genome is in contact with the NL, forming LADs that play important roles in establishing chromosomal structure (van Steensel and Belmont, 2017). While a majority of genes in LADs express at low levels, a subset of them can avoid transcription repression by detaching from the NL (Brueckner et al., 2020). We sought to

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

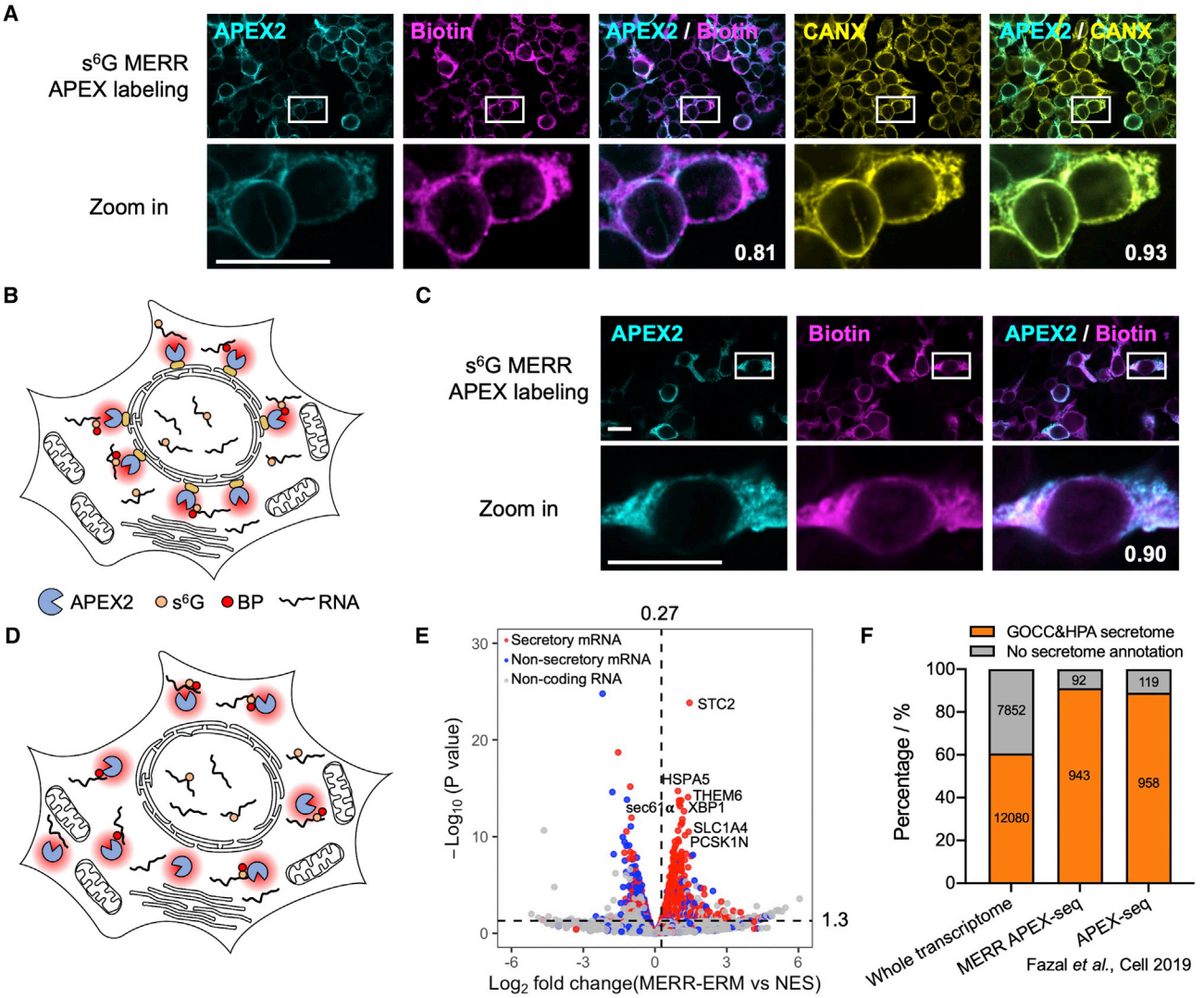


Figure 3. MERR APEX-seq reveals subcellular transcriptome at the ER membrane

(A) Representative immunofluorescence images of HEK293T cells showing (top row, from left to right) APEX2-ER membrane (ERM) expression, biotinylation signal, merged images of APEX2 expression with biotinylation, the ER marker calnexin, and merged images of APEX2 expression with calnexin. Zoomed-in images of boxed regions are shown in the bottom row. Scale bar, 20 μm.

(B) Labeling scheme of s^{6}G MERR APEX-seq at ERM. APEX2 is fused to the N terminus of Sec61β.

(C) Representative images showing (top row, from left to right) APEX2-ERM expression, biotinylation signal, and merged image. Zoomed-in images of boxed regions are shown in the bottom row. Scale bar, 20 μm.

(D) Labeling scheme of s^{6}G MERR APEX-seq in the cytosol. APEX2 is C-terminally fused to a nuclear export sequence (APEX-NES).

(E) Volcano plot of DESeq2 analysis of transcripts enriched by s^{6}G MERR APEX-seq at ERM versus NES. The cutoff p value is chosen as 0.05 (horizontal dotted line) and the cutoff of $\log_2$ fold change of ERM versus NES is set at 0.27 (vertical dotted line).

(F) Comparison of secretory RNA enrichment by MERR APEX-seq and APEX-seq.

See also Figure S3; Data S2 and S3.

understand the spatial organization of newly transcribed RNAs near the NL. Previously, APEX-seq has identified 1,193 transcripts located near lamin A (LMNA), an NL component, in HEK293T cells (Fazal et al., 2019). Notably, LMNA is not strictly located at the NL throughout the cell cycle, while lamin B1 (LMNB1), one of the B-type lamins, is almost exclusively located at the NL (Moir et al., 2000). In this study, we generated a HEK293T cell line stably expressing APEX2 fused to the N terminus of LMNB1 (APEX2-LMNB1) (Figure 4A). Unlike MERR APEX-seq in the mitochondrial matrix, enclosed by two lipid membranes, MERR APEX-seq at the NL is more challenging because NL is an unbounded region in the nucleus, which is easier for a BP radical to diffuse away.

Following metabolic labeling with s^{6}G, cells expressing APEX2-LMNB1 were briefly labeled with 0.5 mM BP for 1 min. Immunofluorescence imaging showed that the biotinylation signal was considerably more diffusive relative to APEX2 expression (Figure 4B). It is likely that biotinylated molecules (both

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

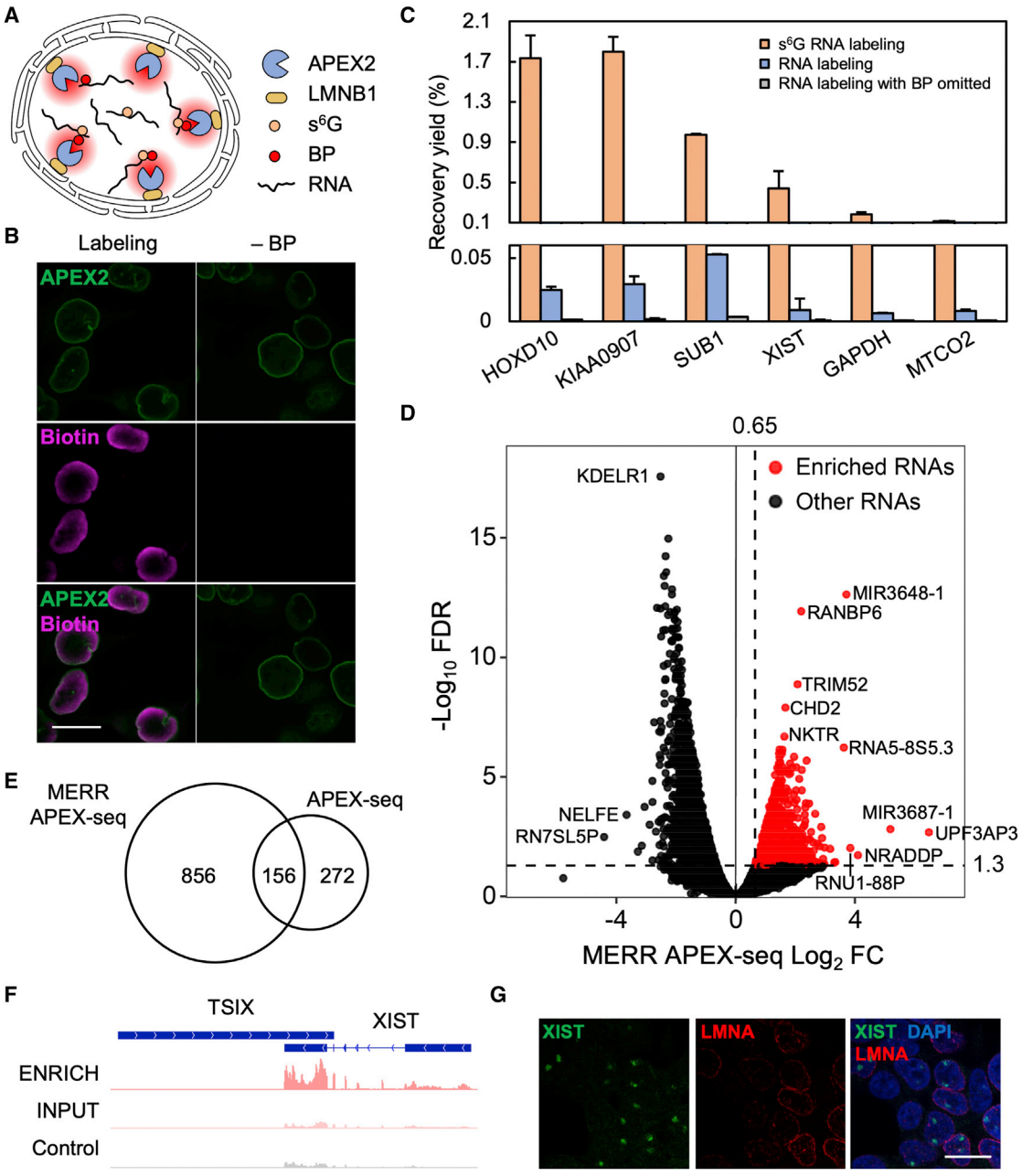


Figure 4. MERR APEX-seq profiling of transcripts at the nuclear lamina

(A) Labeling scheme of s^{6}G MERR APEX-seq at the nuclear lamina (NL). APEX2 is targeted to NL via fusion with LMNB1.

(B) Representative immunofluorescence images of HEK293T cells expressing APEX2-LMNB1 with or without BP labeling. Biotinylation is stained by streptavidin-AF568 (SA-568). Fluorescence intensities in APEX2 and biotin channels have been normalized. Scale bar, 20 μm.

(C) qRT-PCR analysis comparing the recovery rates of several s^{6}G-containing transcripts following APEX-mediated BP labeling at the NL. The cytosol RNA *GAPDH* and the mitochondrial RNA *MTCO2* are the negative RNA markers. The recovery rate is calculated from the C_t values of ENRICH versus INPUT samples for each gene across three technical replicates. Error bars represent SEM.

(D) Volcano plot analysis of transcripts captured by LMNB1 s^{6}G MERR APEX-seq. The false discovery rate (FDR) cutoff is 0.05 (horizontal dotted line). The $\log_2$(ENRICH versus INPUT) cutoff is 0.65 (vertical dotted line).

(E) Venn diagram showing the overlap between LMNB1 s^{6}G MERR APEX-seq and LMNB1 APEX-seq datasets.

(F) Genome tracks of *XIST* and *TSIX* in LMNB1 s^{6}G MERR APEX-seq. BP is omitted from the negative control.

(G) Representative *XIST* smFISH image in HEK293T cells. Scale bar, 20 μm.

See also Figure S4 and Data S4.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

RNAs and proteins) could diffuse into the surrounding nucleoplasm during the 1-min time window of APEX2 labeling. Negative controls omitting the BP probe yielded a negligible biotinylation signal. Biotinylated RNAs were enriched and analyzed by qRT-PCR. The recovery yields for two positive markers, *HOXD10* and *KIAA0907* (defined as $2^{-\Delta Ct}$, where ΔC_t is the difference in C_t values between the ENRICH sample and the INPUT sample), are 1.73% ± 0.23% and 1.80% ± 0.15%, respectively, which were up to 16-fold higher than the negative marker, mt-mRNA *MTCO2* (0.11% ± 0.01%) (Figure 4C). In contrast, for samples omitting s^6G, the recovery yields for *HOXD10* (0.025% ± 0.002%) and *KIAA0907* (0.029% ± 0.006%) were significantly lower. Although the recovery yield for the negative marker *MTCO2* was also reduced to 0.008% ± 0.001%, the overall fold enrichment of *HOXD10* and *KIAA0907* relative to *MTCO2* was still reduced by about 5-fold (Figure S4A).

We performed NGS analysis for MERR APEX-seq at the NL with two biological replicates. We compared MERR APEX-seq at the NL versus mitochondrial matrix (MITO), and MERR APEX-seq versus APEX-seq (i.e., without s^6G metabolic incorporation). Clustering analysis revealed that all MITO labeling methods yield similar sequencing results, while LMNB1 s^6G MERR APEX-seq are different from LMNB1 APEX-seq (Figure S4B). As expected, all INPUT samples were highly correlated (Pearson's correlation coefficient r values from 0.93 to 1.00), whereas the enriched samples in two subcellular regions showed distinct patterns (Figure S4C). We analyzed the differential enriched transcripts using the DESeq2 package in R (Love et al., 2014). The false discovery rate cutoff (adjusted p value) was set to 0.05 and the enrichment yield cutoff, expressed as $\log_2$ (ENRICH versus INPUT), was set to 0.65. These cutoffs yielded a list of 1,012 transcripts in the LMNB1 MERR APEX-seq dataset, including 802 mRNAs (79.2%), 75 antisense RNAs (7.4%), and 38 long intergenic non-coding RNAs (lincRNAs) (3.8%) (Figures 4D and S4D). For comparison, LMNB1 APEX-seq, which was performed in parallel with the only difference being the omission of s^6G metabolic labeling, only enriched 428 RNAs, including 374 mRNAs (87.4%), 12 antisense RNAs (2.8%), and 13 lincRNAs (3.0%) (Figures S4D and S4E). Notably, 156 RNAs (36%) in the LMNB1 APEX-seq were also enriched in LMNB1 MERR APEX-seq (Figure 4E). Thus, MERR APEX-seq is substantially more sensitive than APEX-seq.

We also compared our LMNB1 MERR APEX-seq data with the published LMNA APEX-seq data (Fazal et al., 2019). The coverage of the two methods was similar (1,012 versus 1,193 enriched transcripts) despite 4-fold higher sequencing depth for LMNA APEX-seq (>40 million versus ~10 million paired reads). Among these, 329 transcripts from LMNB1 MERR APEX-seq (33%) overlapped with LMNA APEX-seq. Notably, a substantial portion of the two datasets does not overlap, with 683 transcripts uniquely covered by LMNB1 MERR APEX-seq and 864 by LMNA APEX-seq (Figure S4F). This selectivity may arise from the different choices of protein bait (LMNB1 versus LMNA) and/or the higher sensitivity of MERR APEX-seq toward newly transcribed RNAs. Thus, these two methods could complement each other to improve the overall coverage.

Among the RNAs enriched by MERR APEX-seq at the NL, 778 out of 1,012 (77%) have been annotated as nuclear localized according to the RNALocate database (Zhang et al., 2017), which is consistent with the expectation of NL-proximal labeling (Data S4). Although most genes in LADs are transcriptionally repressed, a subset of genes in LADs has been found to escape the repression (Guelen et al., 2008; Peric-Hupkes et al., 2010). In our LMNB1 MERR APEX-seq dataset, only 33 genes (3%) are located within the LADs (Guelen et al., 2008). Thus, most enriched RNAs are localized to the NL post-transcriptionally (Data S4).

Our dataset also contains several non-coding RNAs that are functionally linked to the NL. Figure 3F shows the enrichment of *XIST*, which helps recruit the inactive X chromosome to the NL (Chen et al., 2016). We performed single-molecule FISH (smFISH) in HEK293T cells to confirm the localization of *XIST* to be NL-proximal (Figure 4G). *MALAT1* and *NEAT1*, which are hallmarks of two closely related nuclear bodies, nuclear speckle and paraspeckle (Clemson et al., 2009; Hutchinson et al., 2007), are also substantially enriched (Figures S4G and S4H). Interestingly, *MALAT1* depletion has been associated with low expression of LMNB1 in human cells (Tripathi et al., 2013). Our data suggest that these two long non-coding RNAs might be related to the NL. Besides, we verified two more transcripts (*EIF3A* and *DDX3X*) uniquely enriched in LMNB1 MERR APEX-seq. The statistics of puncta in smFISH images of both transcripts showed around 47% puncta localized in nucleus, indicating the possibility of both of them anchoring in NL (Figures S4I and S4J).

GOCC analysis of mRNAs enriched by LMNB1 MERR APEX-seq reveals that proteins encoded by these mRNAs are predominantly localized to the nuclear compartments (Figure 5A). Interestingly, LMNA APEX-seq also reported similar findings (Figure S5A) (Fazal et al., 2019). Gene Ontology Biological Process analysis reveals that newly transcribed mRNAs near the NL tend to encode proteins involved in histone modification, chromosomal conformational change, and RNA splicing and processing (Figures 5B, S5B, and S5C). In comparison, LMNA APEX-seq tends to enrich RNAs encoding proteins related to RNA splicing (Figure S5D) (Fazal et al., 2019). Besides, 43 mRNAs encoding nuclear zinc-finger proteins were abundantly enriched in our dataset (Data S4).

Finally, we analyzed the extent of intron retention in the LMNB1 MERR APEX-seq dataset. More intron-retention events were detected in LMNB1 MERR APEX-seq (1,471 events in 832 RNAs) than in LMNB1 APEX-seq (708 events in 409 RNAs) (Figure 5C and Data S5). For example, the *TONSL* mRNA, which encodes a DNA repair protein involved in unmethylated H4K20 incorporation during DNA replication (Saredi et al., 2016), has intron-retention events in LMNB1 MERR APEX-seq (Figure 5D). Kinase implicated in mRNA splicing, CLK3 (Menegay et al., 1999), is an additional example of RNAs with intron-retention events (Figure 5E). When our sequencing datasets were filtered to include only those reads containing hallmarks of nascent RNAs, namely splice junctions and introns, we found that MERR APEX-seq captures substantially more transcripts than APEX-seq (Figure S5E). The higher intron-retention incidence observed in MERR APEX-seq than in APEX-seq may be attributed to its improved sensitivity toward newly transcribed RNAs, which are more likely to exist as splicing intermediates that retain some of their introns. Collectively, these findings demonstrated the power of MERR APEX-seq for profiling newly transcribed RNAs at specific subnuclear compartments.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

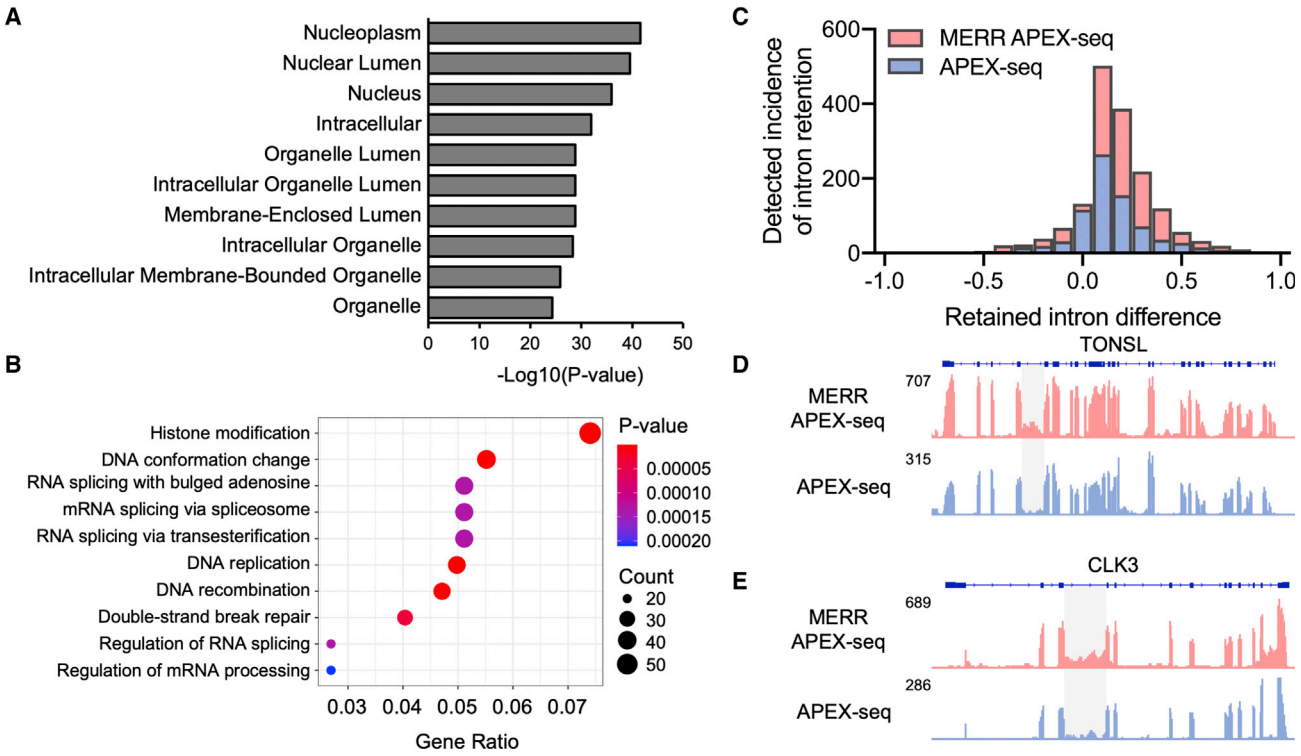


Figure 5. Analysis of nuclear lamina-associated transcriptome

(A and B) GO Cellular Component (A) and GO Biological Process (B) analyses of RNAs enriched in LMNB1 MERR APEX-seq dataset. p values are calculated with Fisher's exact test.

(C) Distributions of detected intron-retention events in MERR APEX-seq and APEX-seq at the nuclear lamina (NL). Retained intron difference is calculated from the normalized reads of transcripts containing retained introns in the ENRICH versus INPUT samples (see STAR Methods).

(D and E) Genome tracks of *TONSL* (D) and *CLK3* (E) in s^6G MERR APEX-seq data reveal intron retention (shaded regions).

See also Figure S5; Data S4 and S5.

MERR APEX-seq tends to identify newly transcribed RNAs

The chemical nature of the incorporated nucleotides underlies the higher labeling efficiency of MERR APEX-seq in comparison with APEX-seq. Our effort to identify the covalent adducts of s^6G-BP and s^4U-BP met with difficulty because of the low reaction yield (Figures S6A–S6E). High-performance liquid chromatography-mass spectrometry suggested additional oxidation of the thiol in s^6G into sulfinic acid (Figure S6C). Recently, Huang et al. (2020) characterized the product of APEX-mediated conjugation of BP with s^4U, which is consistent with our analysis (Figure S6E).

When cells are incubated with s^6G/s^4U for only a short period of time, metabolic incorporation of these nucleobases are restricted to RNAs that are newly transcribed within this time window. We thus performed two sets of experiments to test whether MERR APEX labeling tends to capture newly synthesized transcriptome. In the first set, cells were incubated with s^6G/s^4U for various durations prior to APEX labeling. In mitochondrial matrix, qRT-PCR analysis reveals that APEX labeling is substantially enhanced after incubation with s^6G/s^4U for as short as 0.5 h (Figures 6A and 6B). In the case of ER membrane, the increase of APEX labeling efficiency is delayed for approximately 1 h following incubation with s^6G/s^4U (Figure 6C), which likely arises from the time required for nascent mRNA to mature and translocate from the nucleus to the ER membrane.

In the second set of experiments, cellular transcription was inhibited through addition of actinomycin D (actD) for various periods of time before the initiation of LMNB1 MERR APEX labeling. In the presence of s^6G, RNA recovery yield was reduced by more than (155 ± 52)-fold following 6 h of treatment with actD (Figure 6D). For comparison, the recovery yield was reduced by only (1.6 ± 0.6)-fold in the absence of s^6G incubation (Figures 6E and 6F). Taken together, the above data support the conclusion that MERR APEX-seq preferentially labels newly synthesized RNA molecules.

We noticed that our LMNB1 APEX-seq dataset contained 272 transcripts that were missed by MERR APEX-seq (Figure 4E). To exclude the possibility that metabolic labeling with s^6G might alter mRNA localization, we performed smFISH of three transcripts in this list (*SRSF10*, *ZNF24*, and *ZRANB2*) and quantified the percentage of mRNA puncta localized in the nucleus. No statistically significant difference was observed between untreated cells versus those metabolically labeled with 100 μM s^6G for 5 h, indicating that s^6G incorporation does not influence the subcellular targeting of these transcripts (Figures S6F–S6H). We thus attribute the bias of MERR APEX-seq to its strong preference toward newly synthesized transcripts and transcripts with high GC

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

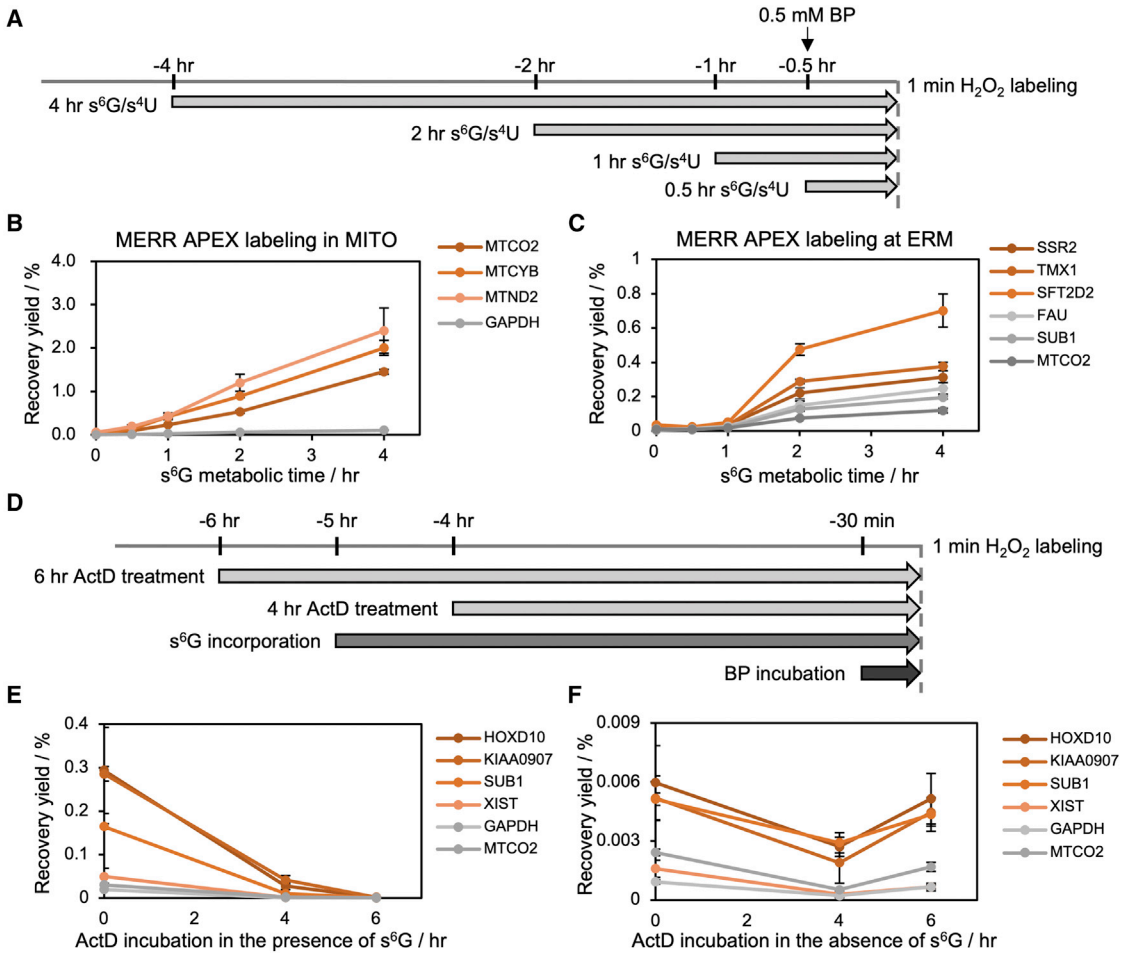


Figure 6. Characterizing the sensitivity of MERR APEX labeling toward newly transcribed RNAs

(A) Experimental scheme of time-course analysis of MERR APEX labeling.

(B and C) qRT-PCR analysis of MITO-MERR APEX2 (B) and MERR APEX2-ERM (C) labeling yields as a function of s^6G incubation time in HEK293T cells.

(D) Experimental scheme of MERR APEX labeling following transcriptional inhibition.

(E and F) qRT-PCR analysis of APEX labeling yields at nuclear lamina following transcriptional inhibition in the presence (E) versus the absence (F) of s^6G.

See also Figure S6 and Data S6.

content, as revealed in our ERM dataset. Therefore, MERR APEX-seq may complement APEX-seq to increase the coverage of RNA profiling at specific subcellular locations. This feature is particularly convenient, since the same APEX2-expressing cell line could be used for both APEX-seq and MERR APEX-seq experiments.

DISCUSSION

Methods to study RNA localization in eukaryotic cells reveal its functions in various biological processes including pre-mRNA splicing, RNA processing, and RNA translation (Buxbaum et al., 2015; Martin and Ephrussi, 2009; Wang et al., 2012). Our new method, MERR APEX-seq, detects newly synthesized RNAs at subcellular resolution with high-throughput sequencing.

In this work, we demonstrated the high spatial specificity of MERR APEX-seq at the ERM. Although APEX2 was positioned to the ERM facing cytosol, which is accessible to all cytoplasmic RNAs, MERR APEX-seq could identify proximal RNAs with high accuracy, capturing a total of 1,035 transcripts with 91% specificity of secretory pathway genes. Interestingly, our datasets also reveal a difference in the enrichment levels toward RNA targets between MERR APEX-seq and conventional APEX-seq (Figure S3F), suggesting that both approaches could complement each other. To understand the cause of this preference, we analyzed the expression levels and nucleobase contents of transcripts captured in our dataset. For expression-level analysis, we used the baseMean value of each transcript as a proxy for its abundance, and compared these values for: (1) transcripts captured by both ERM MERR APEX-seq and ERM APEX-seq datasets; (2) transcripts uniquely captured by ERM MERR APEX-seq; and (3) transcripts uniquely captured by ERM APEX-seq. Our analysis shows that transcripts captured by both methods overall have higher baseMean values than those captured by only one method, indicating that enrichment of lowly expressed transcripts are more sensitive to the preference of

labeling methods (Figure S3I). For nucleobase content analysis, our data show that transcripts enriched in s^6G MERR APEX-seq overall have higher GC content than transcripts enriched in APEX-seq, suggesting that MERR APEX-seq is biased toward transcripts containing more Gs, as compared with APEX-seq (Figure S3G). It is also possible that certain transcripts are missed by MERR APEX-seq due to their slower turnover rate, which leads to fewer copies of nascent transcripts.

We have applied MERR APEX-seq to study the local transcriptome of the NL. While both LMNA and LMNB1 are NL markers, only LMNA has been previously employed as bait in APEX-seq experiments (Fazal et al., 2019). Unlike LMNA, which is diffusively localized throughout the nucleoplasm in early G_1 phase in mammalian cells, LMNB1 is exclusively localized at the nuclear rim (Moir et al., 2000). Super-resolution microscopy with three-dimensional structured illumination microscopy shows that LMNA and LMNB1 are organized into distinct supramolecular structures (Shimi et al., 2015). Considering the differential distribution of LMNA and LMNB1, we chose LMNB1 as the bait for NL MERR APEX-seq. Our dataset primarily enriches transcripts encoding proteins involved in histone remodeling, suggesting a potential functional connection between lamina-proximal transcription and chromatin remodeling.

A potential limitation of the method is the cytotoxicity of prolonged metabolic labeling. While non-canonical ribonucleosides have been widely used as tools for biological research (Eremeeva and Herdewijn, 2019), incubating cultured cells with over 50 μM s^4U has been reported to inhibit production and processing of 47S rRNA (Burger et al., 2013). Similarly, treatment with over 100 μM s^6G could inhibit rRNA maturation (Weiss and Pitot, 1974). Thus, MERR APEX-seq should be avoided in studying ribosome biology.

We envision that MERR APEX-seq could be applied to expand our knowledge of RNA dynamics, localizations, and functions in many more subcellular regions, particularly those are not easily accessible via conventional fractionation techniques, including neuronal synapse and growth cone. The higher sensitivity exhibited by MERR APEX-seq compared with APEX-seq makes it more feasible to use in profiling of spatial transcriptomes with limited biomaterials. The incorporation of non-canonical nucleosides also makes it possible to study the subcellular response in the time dimension.

SIGNIFICANCE

In this article, we present a proximity-dependent nascent RNA labeling technique called MERR APEX-seq. MERR APEX-seq combines the metabolic incorporation of electron-rich ribonucleosides, 6-thioguanosine and 4-thiouridine, with the peroxidase-mediated RNA labeling method, APEX-seq. MERR APEX-seq achieves more than 20-fold higher labeling efficiency than APEX-seq in mitochondrial matrix and offers high spatial specificity and high sensitivity at ER of live cells. MERR APEX-seq unveils an inventory of nuclear lamina-proximal transcriptome that contains more than 1,000 RNAs, many of which encode for nuclear functional proteins. MERR APEX-seq could thus expand our knowledge of RNA localization and function in subcellular compartments.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mammalian cell culture
  - Generation of HEK293T cells stably expressing different APEX2 constructs
- METHOD DETAILS
  - Protein expression and purification
  - LC-MS characterization of RNA labeling
  - LC-MS/MS analysis of s^6G/s^4U metabolic rate
  - Dot blot analysis of *in vitro* RNA labeling
  - RNA biotinylation in living cells
  - Immunofluorescence microscopy
  - Total RNA extraction
  - DNA digestion and quality control of INPUT samples
  - Affinity purification of biotinylated RNA
  - qRT-PCR analysis
  - Time course assay
  - Transcriptional inhibition assay
  - Next generation sequencing
  - Single-molecule FISH imaging
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - NGS data analysis
  - NGS data visualization
  - Statistical methods

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chembiol.2022.02.005.

ACKNOWLEDGMENTS

We thank Y. Zhou and G. Wang for assistance with plasmid construction, P. Wang and G. Wang for advice on smFISH, L. Peng for assistance with image analysis, T. Yue for assistance with biotin-phenol probe synthesis, and W. Tang, T. Ding, Y. Fang, and Y. Li for helpful discussions. We thank Prof. Jing Wang (Peking University, China) for providing s^4U oligonucleotides. This work was supported by the Ministry of Science and Technology (2018YFA0507600, 2017YFA0503600) and the National Natural Science Foundation of China (32088101, 91753131, 21727806). P.Z. is sponsored by Li Ge-Zhao Ning Life Science Junior Research Fellowship and Bayer Investigator Award. The authors thank the National Center for Protein Sciences at Peking University in Beijing, China, for assistance with Fragment Analyzer 12. The measurement of nuclear magnetic resonance was performed at the Analytical Instrumentation Center of Peking University.

AUTHOR CONTRIBUTIONS

P.Z. conceived the project. R.L., Z.Z., and P.Z. designed experiments. R.L. and W.W. performed all experiments. R.L. and P.Z. analyzed data. R.L., Z.Z., and P.Z. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 13, 2020
Revised: July 30, 2021
Accepted: February 10, 2022
Published: March 3, 2022

REFERENCES

Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell *154*, 914–927.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arias, Z.G., Alvarez, J.L.M., and Fonseca, J.M.L. (2006). Formation and electrochemical characterization of 6-thioguanosine monolayers on the mercury surface. J. Colloid Interf. Sci. *300*, 60–68.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Brueckner, L., Zhao, P.A., van Schaik, T., Leemans, C., Sima, J., Peric-Hupkes, D., Gilbert, D.M., and van Steensel, B. (2020). Local rewiring of genome-nuclear lamina interactions by transcription. EMBO J. *39*, e103159.

Burger, K., Muhl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dolken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. *10*, 1623–1630.

Buxbaum, A.R., Haimovich, G., and Singer, R.H. (2015). In the right place at the right time: visualizing and understanding mRNA localization. Nat. Rev. Mol. Cell Biol. *16*, 95–109.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Ami, G.O.H., and Web Presence Working, G. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289.

Carlson, M. (2020). org.Hs.eg.db: Genome Wide Annotation for Human (Bioconductor). https://bioconductor.org/packages/org.Hs.eg.db/.

Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science *354*, 468–472.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol. Cell *33*, 717–726.

Day-Richter, J., Harris, M.A., Haendel, M., Gene Ontology OBO-Edit Working Group, and Lewis, S. (2007). OBO-Edit–an ontology editor for biologists. Bioinformatics *23*, 2198–2200.

Dobrzynska, A., Gonzalo, S., Shanahan, C., and Askjaer, P. (2016). The nuclear lamina in health and disease. Nucleus *7*, 233–248.

Doerr, I.L., Wempen, I., Clarke, D.A., and Fox, J.J. (1961). Thiation of nucleosides. III. Oxidation of 6-Mercaptopurines1. J. Org. Chem. *26*, 3401–3409.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat. Rev. Mol. Cell Biol. *17*, 756–770.

Eremeeva, E., and Herdewijn, P. (2019). Reprint of: non canonical genetic material. Curr. Opin. Biotechnol. *60*, 259–267.

Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of subcellular RNA localization revealed by APEX-seq. Cell *178*, 473–490 e426.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

Hansen, K.D., Gentry, J., Long, L., Gentleman, R., Falcon, S., Hahne, F., and Sarkar, D. (2020). Rgraphviz: Provides Plotting Capabilities for R Graph Objects (R package version 2.8.1. 2009).

Heihoff, K., Redmond, R.W., Braslavsky, S.E., Rougee, M., Salet, C., Favre, A., and Bensasson, R.V. (1990). Quantum yields of triplet and O-2(1-Delta-G) formation of 4-thiouridine in water and acetonitrile. Photochem. Photobiol. *51*, 635–641.

Holzer, K.P., and Wrona, M.Z. (1983). Electrochemical oxidation of 4-thiouracil to Bis(4-Thiouracil)Disulfide and chemical-transformations of the disulfide. Bioelectrochem. Bioener *10*, 199–212.

Huang, J., Zhao, R., Qin, S., Yang, S., Li, W., Mo, J., Wang, F., Weng, X., and Zhou, X. (2020). 4-Thiouridine enhanced peroxidase-generated biotinylation of RNA. Chembiochem. *22*, 212–216.

Hubstenberger, A., Courel, M., Benard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.B., Munier, A., Fradet, M., et al. (2017). P-body purification reveals the condensation of repressed mRNA regulons. Mol. Cell *68*, 144–157.e5.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics *8*, 39.

Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science *346*, 1257521.

Kaewsapsak, P., Shechner, D.M., Mallard, W., Rinn, J.L., and Ting, A.Y. (2017). Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. Elife *6*, e29224.

Kall, L., Krogh, A., and Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC table browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

Khong, A., Matheny, T., Jain, S., Mitchell, S.F., Wheeler, J.R., and Parker, R. (2017). The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. Mol. Cell *68*, 808–820.e5.

Kiefer, L., Schofield, J.A., and Simon, M.D. (2018). Expanding the nucleoside recoding toolkit: revealing RNA population dynamics with 6-thioguanosine. J. Am. Chem. Soc. *140*, 14567–14570.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, E., and Jung, H. (2015). Local protein synthesis in neuronal axons: why and how we study. BMB Rep. *48*, 139–146.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. Cell *153*, 178–192.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. *8*, 761–773.

Krijthe, J.H.. Rtsne: T-distributed stochastic neighbor embedding using a Barnes-hut implementation. https://github.com/jkrijthe/Rtsne.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Lefebvre, F.A., Cody, N.A.L., Bouvrette, L.P.B., Bergalet, J., Wang, X., and Lecuyer, E. (2017). CeFra-seq: systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing. Methods *126*, 138–148.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., and Yi, C. (2016). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. Nat. Chem. Biol. *12*, 311–316.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell *136*, 719–730.

Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J., and Keir, H.M. (1978). Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur. J. Biochem. *92*, 373–379.

Menegay, H., Moeslein, F., and Landreth, G. (1999). The dual specificity protein kinase CLK3 is abundantly expressed in mature mouse spermatozoa. Exp. Cell Res. *253*, 463–473.

Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.M., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., et al. (2011). The human mitochondrial transcriptome. Cell *146*, 645–658.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47*, D419–D426.

Mofatteh, M., and Bullock, S.L. (2017). SnapShot: subcellular mRNA localization. Cell *169*, 178–178 e171.

Moir, R.D., Yoon, M., Khuon, S., and Goldman, R.D. (2000). Nuclear lamins A and B1: different pathways of assembly during nuclear envelope formation in living cells. J. Cell Biol. *151*, 1155–1168.

Padiath, Q.S. (2019). Autosomal dominant leukodystrophy: a disease of the nuclear lamina. Front. Cell Dev. Biol. *7*, 41.

Padron, A., Iwasaki, S., and Ingolia, N.T. (2019). Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. Mol. Cell *75*, 875–887 e875.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell *38*, 603–613.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Raj, A., and Tyagi, S. (2010). Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. Methods Enzymol. *472*, 365–386.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Robinson, J.T., Thorvaldsdottir, H., Wenger, A.M., Zehir, A., and Mesirov, J.P. (2017). Variant review with the integrative genomics viewer. Cancer Res. *77*, e31–e34.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Saredi, G., Huang, H., Hammond, C.M., Alabert, C., Bekker-Jensen, S., Forne, I., Reveron-Gomez, N., Foster, B.M., Mlejnkova, L., Bartke, T., et al. (2016). H4K20me0 marks post-replicative chromatin and recruits the TONSL-MMS22L DNA repair complex. Nature *534*, 714–718.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods *15*, 221–225.

Secrist, J.A., 3rd, Barrio, J.R., and Leonard, N.J. (1971). Attachment of a fluorescent label to 4-thiouracil and 4-thiouridine. Biochem. Biophys. Res. Commun. *45*, 1262–1270.

Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. U S A. *111*, E5593–E5601.

Shimi, T., Kittisopikul, M., Tran, J., Goldman, A.E., Adam, S.A., Zheng, Y., Jaqaman, K., and Goldman, R.D. (2015). Structural organization of nuclear lamins A, C, B1, and B2 revealed by superresolution microscopy. Mol. Biol. Cell *26*, 4075–4086.

Steenken, S., and Jovanovic, S.V. (1997). How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. J. Am. Chem. Soc. *119*, 617–618.

Tang, C.W., Maya-Mendoza, A., Martin, C., Zeng, K., Chen, S., Feret, D., Wilson, S.A., and Jackson, D.A. (2008). The integrity of a lamin-B1-dependent nucleoskeleton is a fundamental determinant of RNA synthesis in human cells. J. Cell Sci. *121*, 1014–1024.

The Gene Ontology, C. (2019). The gene ontology resource: 20 years and still going strong. Nucleic Acids Res. *47*, D330–D338.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. *14*, 178–192.

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S.M., Wu, X., Zhang, Y., Gorospe, M., Prasanth, S.G., Lal, A., et al. (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet *9*, e1003368.

Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M., et al. (2016). smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. *44*, e165.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

van der Maaten, L.J.P., and Hinton, G.E. (2008). Visualizing data using t-SNE. J. Mach Learn. Res. *9*, 2579–2605.

van Steensel, B., and Belmont, A.S. (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. Cell *169*, 780–791.

Wang, E.T., Cody, N.A., Jog, S., Biancolella, M., Wang, T.T., Treacy, D.J., Luo, S., Schroth, G.P., Housman, D.E., Reddy, S., et al. (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell *150*, 710–724.

Wang, P., Tang, W., Li, Z., Zou, Z., Zhou, Y., Li, R., Xiong, T., Wang, J., and Zou, P. (2019). Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. Nat. Chem. Biol. *15*, 1110–1119.

Weiss, J.W., and Pitot, H.C. (1974). Inhibition of ribosomal RNA maturation in Novikoff hepatoma cells by toyocamycin, tubercidin, and 6-thioguanosine. Cancer Res. *34*, 581–587.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

Wickham, H. (2019). Stringr: Simple, Consistent Wrappers for Common String Operations (R package version 1.4.0.). https://CRAN.R-project.org/package=stringr.

Wilkie, G.S., and Schirmer, E.C. (2008). Purification of nuclei and preparation of nuclear envelopes from skeletal muscle. Methods Mol. Biol. *463*, 23–41.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic Acids Res. *48*, D682–D688.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Yu, Y., Zhang, M., Liu, J., Xu, B., Yang, J., Wang, N., Yan, S., Wang, F., He, X., Ji, G., et al. (2018). Long non-coding RNA PVT1 promotes cell proliferation and migration by silencing ANGPTL4 expression in cholangiocarcinoma. Mol. Ther. Nucleic Acids *13*, 503–513.

Zhang, T., Tan, P., Wang, L., Jin, N., Li, Y., Zhang, L., Yang, H., Hu, Z., Zhang, L., Hu, C., et al. (2017). RNALocate: a resource for RNA subcellular localizations. Nucleic Acids Res. *45*, D135–D138.

Zhang, Z. (2016). Reshaping and aggregating data: an introduction to reshape package. Ann. Transl Med. *4*, 78.

Zhou, X., Li, D., Zhang, B., Lowdon, R.F., Rockweiler, N.B., Sears, R.L., Madden, P.A., Smirnov, I., Costello, J.F., and Wang, T. (2015). Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. Nat. Biotechnol. *33*, 345–346.

Zhou, X., Maricque, B., Xie, M., Li, D., Sundaram, V., Martin, E.A., Koebbe, B.C., Nielsen, C., Hirst, M., Farnham, P., et al. (2011). The human Epigenome browser at Washington University. Nat. Methods *8*, 989–990.

Zhou, Y., Wang, G., Wang, P., Li, Z., Yue, T., Wang, J., and Zou, P. (2019). Expanding APEX2 substrates for proximity-dependent labeling of nucleic acids and proteins in living cells. Angew. Chem. Int. Ed. Engl. *58*, 11763–11767.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-V5	Biodragon	Cat# B1005F
Goat anti-mouse-Alexa Fluor 488	ThermoFisher	Cat# A-11029; RRID: AB_2534088
Goat anti-rabbit-Alexa Fluor 568	ThermoFisher	Cat# A-11036; RRID: AB_10563566
Streptavidin-Alexa Fluor 647	ThermoFisher	Cat# S21374; RRID: AB_2336066
Streptavidin-HRP	Pierce	Cat# 21124
Rabbit anti-Calnexin	Abcam	Cat# ab22595; RRID: AB_2069006
Rabbit anti-lamin A	Abcam	Cat# ab26300; RRID: AB_775965
Chemicals, peptides, and recombinant proteins		
Tryptone	OXOID	Cat# CM0129B
Yeast extract	OXOID	Cat# LP0021
NaCl	Sigma	Cat# S3014
Ampicillin	Inalco	Cat# 1758-9314
$Na_2HPO_4 \cdot 12H_2O$	Tongguang	Cat# 107035
KH_2PO_4	Tongguang	Cat# 117013
Glycerol	Xilong	Cat# 70-20
Glucose	Sigma	Cat# G6152
Galactose	Sinopharm	Cat# 63004434
Tris-HCl	Aladdin	Cat# T105287
Ni-NTA agarose	Qiagen	Cat# 30210
DMEM	Gibco	Cat# C11995500BT
Trypsin	Life	Cat# 25200056
Fetal bovine serum	Gibco	Cat# 10099141
LIPO-2000	Invitrogen	Cat# 11668019
Opti-MEM	Gibco	Cat# 31985062
Blasticidin	Selleck	Cat# S7419
4-thiouridine	Sigma	Cat# T4509
6-thioguanosine	Sigma	Cat# 858412
TRIzol reagent	Invitrogen	Cat# 15596018
Sodium ascorbate	Aladdin	Cat# S105024
6-Hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox)	Sigma	Cat# 238813
DMSO	Sigma	Cat# D5879
1 M Tris-HCl buffer, pH 7.5	Invitrogen	Cat# 15567027
5 M NaCl, RNase free	Ambion	Cat# AM9759
Tween-20	Solarbio	Cat# T8200
Ultrapure water	Beyotime	Cat# ST872
Matrigel matrix	Corning	Cat# 356234
PBS	Solarbio	Cat# P1020
Paraformaldehyde	Sinopharm	Cat# 80096692
Triton X-100	Sigma	Cat# T8787
BSA	Sangon	Cat# A500023-0100
DAPI	ThermoFisher	Cat# D1306
DNase I	NEB	Cat# M0303
0.5 M EDTA, pH 8.0	Amresco	Cat# E522

(*Continued on next page*)

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
1N NaOH	Sigma	Cat# S2770
Yeast tRNA	Gibco	Cat# 15401011
Glycogen, RNA grade	Fermentas	Cat# R0551
RNase-free PBS	Life	Cat# AM9624
Formamide	Sigma	Cat# F9037
D-biotin	Invitrogen	Cat# B20656
Chloroform	Tongguang	Cat# 112049
Isopropanol	Tongguang	Cat# 106030
SYBR Green Master Mix	Life	Cat# A25742
10×NEB3.1 buffer	NEB	Cat# B7203V
20×SSC buffer	Amresco	Cat# 0804
Vanadyl ribonucleoside complex (VRC)	Sigma	Cat# 94740
Dextran sulfate sodium salt	Sigma	Cat# D6001
Actinomycin D	GlpBio	Cat# GC16866
Nuclease P1	Sigma	Cat# N8630
Alkaline phosphatase	Sigma	Cat# P4252
MES buffer, pH 6.5	Jena Bioscience	Cat# BU-109
VAHTS DNA Clean Beads	Vazyme	Cat# N411-02
Critical commercial assays		
RNA Clean and concentrator-25	Zymo Research	Cat# R1018
SuperScript III Reverse Transcriptase	Invitrogen	Cat# 18080044
NEBNext Ultra II RNA Library Prep Kit for Illumina	NEB	Cat# E7770
NEBNext® Poly(A) mRNA Magnetic Isolation Module	NEB	Cat# E7490
Deposited data		
Raw and analyzed data	This paper	GEO: GSE192739
HEK293 ER proximity-specific ribosome profiling RNA-seq	Jan et al. (2014)	PMID: 25378630
Human LADs	Guelen et al. (2008)	PMID: 18463634
GO enrichment analysis	Mi et al. (2019)	http://www.pantherdb.org/
Signal peptide annotations (SignalP 4.0)	Petersen et al. (2011)	PMID: 21959131
Signal peptide annotations (Phobius)	Kall et al. (2004)	PMID: 15111065
Transmembrane protein annotations (TMHMM)	Krogh et al. (2001)	http://www.cbs.dtu.dk/services/TMHMM/
Experimental models: Cell lines		
HEK293T	Laboratory of Prof. Jing Yang, Peking University	N/A
Oligonucleotides		
Primers used	This study	Data S6
smFISH oligos	This study	Data S6
Recombinant DNA		
Flag-APEX2	This study	EcoR1-FRB-Flag-APEX2-XhoI-6xHis
MITO-APEX2	This study	BstBI-MITO-V5-APEX2-NheI MITO: LATRVFSLVGKRAISTSVCVRAH
APEX2-ERM	This study	BstBI-V5-APEX2-sec61β-NheI
APEX2-NES	This study	BamHI-V5-APEX2-NES-NheI NES: LQLPPLERLTLD
APEX2-LMNB1	This study	BstBI-V5-AgeI-APEX2-BamHI-LMNB1-NheI
Software and algorithms		
MultiQC package	Ewels et al. (2016)	RRID: SCR_014982
HISAT2	Kim et al. (2015)	RRID:SCR_015530

(*Continued on next page*)

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HTSeq	Anders et al. (2015)	RRID: SCR_005514
Bioconductor	Gentleman et al. (2004)	RRID: SCR_006442
DESeq2	Love et al. (2014)	RRID: SCR_015687
Ggplot2	Wickham (2016)	RRID: SCR_014601
t-SNE	van der Maaten and Hinton (2008)	RRID: SCR_016900
SAMtools	Li et al. (2009)	PMID: 19505943
Bedtools	Quinlan and Hall (2010)	RRID: SCR_006646
Deeptools	Ramirez et al. (2014)	RRID:SCR_016366
IGV	Robinson et al. (2011)	RRID:SCR_011793
WashU Epigenome Browser	Zhou et al. (2011)	PMID: 22127213
rMATS	Shen et al. (2014)	RRID: SCR_013049

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Peng Zou (zoupeng@pku.edu.cn).

Materials availability

All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Data and code availability

All data presented are available in the main text and supplemental materials. The accession number for the raw sequencing data reported in this paper is Gene Expression Omnibus (GEO): GSE192739. All original code is available in this paper's supplemental information. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mammalian cell culture

HEK293T cells with or without APEX2 expression (passages < 20) were cultured in complete medium consisting of Dulbecco's Modified Eagle's Medium (DMEM, Gibco, C11995500BT) supplemented with 10% fetal bovine serum (Gibco, 10099141) at 37 °C with 5% CO_2. For *in vitro* RNA dot blot experiments, qPCR experiments and RNA-seq experiments, cells were plated into 6-well plates (Corning). For immunofluorescence microscopy imaging experiments, cells were plated on glass coverslips into 24-well plates.

Generation of HEK293T cells stably expressing different APEX2 constructs

For lentivirus preparation, HEK293T cells (passages < 10) were seeded into 6-well plates and transfected at approximately 60% confluency with the lentiviral vector pLX304 fusing APEX2 to located sequences (2 μg) and two packaging plasmids pVSVG (1.4 μg) and dR8.9 (2 μg) with the help of 10.8 μL LIPO-2000 (Invitrogen, 11668019) in 160 μL Opti-MEM (Gibco, 31985062) for each well. The mixture was carefully added to the cells cultured in 6-well plate after replacing complete medium to 2 mL DMEM. Following incubation at 37 °C for 4 hr, the medium was changed to complete medium. The lentivirus was collected after 36-48 hr and filtered through 0.45 μm filter membrane to remove particulates and cell debris. The lentivirus was aliquoted and quick-frozen in liquid nitrogen, then stored at −80°C. HEK293T cells (passages < 10) at approximately 70% confluency were infected by lentivirus with a proper titration for 48 hr. Cells were selected by 5 μg/mL blasticidin (Selleck, S7419) in complete medium for 7 days. APEX2 expression in these cells were verified by V5-tag immunofluorescence before further experiments.

METHOD DETAILS

Protein expression and purification

Escherichia coli BL21 (DE3) competent cells were transformed with 140 ng pET21a-APEX2-His_6 plasmid. Transformed bacteria were cultured on LB-agar medium without antibiotics at 37°C overnight. A single colony was picked to inoculate 5 mL LB medium (10 g/L Tryptone, 5 g/L Yeast Extract and 10 g/L NaCl) containing 100 μg/mL ampicillin and incubated at 37°C for 8 hr. The bacterial culture was diluted by 1:100 into 250 mL auto-induction SB medium (25 mM $Na_2HPO_4 \cdot 12H_2O$, 25 mM KH_2PO_4, 20 g/L Tryptone, 5 g/L Yeast

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

Extract and 86 mM NaCl) with 10 mL sugar mix (150 mL glycerol, 12.5 g glucose and 50 g galactose in 1 L ddH_2O). Protein expression was induced at 20°C for 40 hr. Cells were collected by centrifugation at 7,000 rpm for 25 min. The cell pellet was resuspended in the binding buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl) and lysed by ultrasonication (Sonics Uibra cell) on ice for 40 min at 30% max intensity. Following centrifugation at 10,000 rpm for 30 min at 4°C, the supernatant was collected to incubate with 1.5 mL Ni-NTA agarose beads (Qiagen, 30210) at 4°C for 30 min with rotation. The Ni-NTA agarose beads with protein were loaded into a column and washed with the binding buffer containing 10 mM imidazole to remove non-specifically bound proteins. The protein was eluted by elution buffer (binding buffer containing 500 mM imidazole). Excessive imidazole was removed through dialysis against PBS at 4°C for 24 hr. The purified protein was concentrated by ultrafiltration (Amicon Ultra-4 centrifugal filter, EMD Millipore, UFC801096). The protein was aliquoted and stored at −80°C.

LC-MS characterization of RNA labeling

0.5 mM s^4U (in ddH_2O) or s^6G (in DMSO), 2 mM BP (in DMSO) and 5 μM APEX2 were mixed in PBS. 1 mM H_2O_2 (final concentration) was added to the reaction mixture for labeling reaction and sample with H_2O_2 omitted was performed as negative control. After 1 min at room temperature, the reaction was diluted by 20 times with ddH_2O and filtered by 0.22 μm MCE membrane filter (Navigator, NMF04–2). The dilution was 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) with a Waters C18 Sun Fire separation column (5 μm, 150 × 4.6 mm^2). The m/z values of nucleotides and products were collected with negative ion mode. The m/z values of BP were collected with positive ion mode. Extracted ions were identified and integrated using MassLynx Mass Spectrometry Software. The adducts of BP with s^6G (m/z = 694) and s^4U (m/z = 621) were recorded (Figures S6A–S6E). UV spectrum of both s^6G and s^4U MERR APEX-seq labeling at a wavelength of 310 nm was also recorded with new peaks at the retention time of adduct peaks (Doerr et al., 1961; Secrist et al., 1971).

Time (min)	Flow Rate (mL/min)	0.1% Formic acid (%)	Methanol (%)
0	0.300	98.0	2.0
1.00	0.300	98.0	2.0
5.50	0.300	10.0	90.0
7.00	0.300	2.0	98.0
8.00	0.300	2.0	98.0

LC-MS/MS analysis of s^6G/s^4U metabolic rate

For mass spectrometry analysis of non-canonical ribonucleoside incorporation rate, total cellular RNA was extracted from HEK293T cells following 0-4 hr incubation with 100 μM s^6G or s^4U. After DNase I digestion and RNA purification, 200 ng RNA for each sample was digested into nucleosides by 0.5 U nuclease P1 (Sigma, N8630) in 20 μL buffer containing 10 mM ammonium acetate, pH 5.3 at 42°C for 6 hr. The mixture was then added with 0.5 U alkaline phosphatase (Sigma, P4252) in 1.25 μL 1 M MES buffer, pH 6.5 (Jena Bioscience, BU-109) and reacted at 37°C for 6 hr (Li et al., 2016). The reaction was diluted to 50 μL and filtered with 0.22 μm MCE membrane filter. Thereafter, 5 μL of the solution was injected into LC-MS/MS equipped with a C18 column UPLC and triple-quadrupole mass spectrometry (AB SCIEX QTRAP 5500). The mass shift of m/z 300 to 168 (s^6G) and 261 to 129 (s^4U) were monitored in the positive ion multiple reaction-monitoring (MRM) mode. A series of standard s^6G and s^4U (from 5 nM to 1 μM) was injected as external standards for quantitation. Concentrations of nucleosides in extracted RNA samples were deduced by fitting the signal intensities into the standard curves. The ratios of s^6G/G and s^4U/U were subsequently calculated.

Dot blot analysis of *in vitro* RNA labeling

HEK293T cells were metabolized by final concentration of 100 μM s^4U (Sigma, T4509) overnight at approximately 60% confluency or by 100 μM s^6G (Sigma, 858412) for 5 hr at approximately 70% confluency or by none. Total RNA was extracted by TRIzol Reagent (Invitrogen, 15596018) following the manufacturer's instructions. For each labeling sample, taking 5 μg RNA reacted with 5 μM purified APEX2, 500 μM (final concentration) BP probe dissolved in DMSO, and 1 mM (final concentration) H_2O_2 for 1 min at room temperature. Reactions were quenched by 10 mM sodium ascorbate, 5 mM Trolox and 10 mM NaN_3. Then small molecules and proteins were removed by RNA Clean & Concentrator kit (Zymo, R1015). The same amount of purified RNA for each sample were loaded onto Immobilon-Ny+ membrane (Merck Millipore, INYC00010-1) and crosslinked to the membrane under 254 nm ultraviolet by an ultraviolet crosslinker (Analytik Jena). The membrane was blocked by 3% BSA in TBST (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Tween-20 (Solarbio, T8200)) at room temperature for 1 hr with gentle shake. Then the membrane was incubated with streptavidin-HRP (Pierce, 21124) by 1:4,000 dilution for 1 hr. After 3 times washing by TBST for 5 min each time, the membrane was incubated with Clarity Western ECL Substrate (Bio-Rad, 1705061) and imaged through a ChemiDoc imaging system (Bio-Rad). For Figure S3B, there were 3 kinds of RNA were used. For labeling and omitting BP samples, 23 nt RNA oligonucleotides with one s^4U incorporation (23-mer) was used for labeling. For omitting s^4U samples, RNA extracted from HEK293T cells was used for labeling. Standard

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

biotin-RNA was 80 nt oligonucleotides with one biotin modification. 5 μg RNA was incubated with 5 μM APEX2, 0.5 mM BP and 1 mM H_2O_2 at room temperature for 1 min. Each sample dot was dropped with 150 ng RNA, and standard biotin-RNA was mixed with total RNA up to 150 ng. Relative signal intensity and weight of standard biotin-RNA were fitted the linear regression equation of $y = 12482x - 5283$. Relative signal intensity of labeling and omitting s^4U sample were 14994 and 342, respectively. Then, s^4U-biotin intensity contained 150 ng labeling RNA was equivalent to standard biotin-RNA weight of $\frac{14994-343+5283}{12482} = 1.60\,ng$. The s^4U-biotin intensity equal to $\frac{1.60\ ng*10^{-9}*6.02*10^{23}}{80\ nt*320g/mol} = 3.76*10^{10}$ biotin molecules. And in 150 ng labeling sample, there were $\frac{150\ ng*10^{-9}*6.02*10^{23}}{23\ nt*320\ g/mol} = 1.23*10^{13}$ s^4U molecules. Thus, each biotin molecule was added to $\frac{1.23*10^{13}}{3.76*10^{10}} = 326$ s^4U molecules.

RNA biotinylation in living cells

HEK293T cells were grown to approximately 60% confluency and s^4U or s^6G were added to the medium to a final concentration of 100 μM and incubated at 37°C for 4.5 hr in the dark. Then 500 μM BP probe was added to medium directly for 30 min. 1 mM H_2O_2 was added to each well in the dark and the plate was agitated for 1 min. The reaction was quenched 3 times by replacing the medium with an equal volume of 10 mM sodium ascorbate, 5 mM Trolox and 10 mM NaN_3 in PBS.

Immunofluorescence microscopy

After RNA labeling, cells were washed twice by PBS to get rid of the free small molecules. Cells were fixed by 4% paraformaldehyde in PBS at room temperature for 15 min and washed twice by PBS to remove extra paraformaldehyde. Then, cells were permeabilized by 0.1% Triton X-100 (Sigma, T8787) in PBS for 15 min followed by 3 times washing. Blocking buffer (3% BSA in PBS) was used to block cells at room temperature for 30 min with gentle shake. Cells were incubated with primary antibody (Mouse-anti-V5 antibody, 1:1,000 dilution) in blocking buffer at room temperature for 1 hr. After being washed by blocking buffer with 0.05% Tween-20 for 3 times, cells were incubated by secondary antibody (Alexa Fluor goat anti-mouse-488, ThermoFisher, 1:1,000 dilution) and Streptavidin-Alexa Fluor 647 (ThermoFisher, 1:1,000 dilution) or Streptavidin-Alexa Fluor 568 (ThermoFisher, 1:1,000 dilution) in blocking buffer at room temperature for 1 hr with gentle rotation. Following washed by blocking buffer with 0.05% Tween-20 for 3 times, cells were counterstained with DAPI (ThermoFisher, D1306) in PBS at room temperature for 15 min. Then cells were washed 3 times by PBS before microscopy imaging.

The Immunofluorescence microscopy imaging platform was based on an inverted fluorescence microscope (Nikon-TiE) 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 platform was controlled by a customized software written in LabVIEW v.15.0 (National Instruments). In this study, we used 60× oil immersion-lens to get images. Image analysis was performed on ImageJ (Schindelin et al., 2012; Schneider et al., 2012).

Total RNA extraction

To MERR APEX labeled or unlabeled (control) HEK293T cells in each well of 6-well plate, we discarded quenching buffer and added 1 mL TRIzol reagent for cell lysis. After pipetting to homogenize, the cell lysate was transferred to 1.5 mL Eppendorf tube. 0.2 mL chloroform were added to each tube then shake the tube several times with lid close. The samples were incubated for 5 min at room temperature and centrifuged for 15 min at 12, 000 × g at 4°C. We transferred the aqueous phase containing RNA to a new tube and added 0.5 mL isopropanol to mix well. The mixture was incubated at −20°C for 20 min then centrifuged for 15 min at 12, 000 × g at 4°C to get pellet at the bottom of the tube. After discarding the supernatant, the pellet was washed by 1 mL 75% ethanol with vortex. Samples were centrifuged for 5 min at 12, 000 × g at 4°C then supernatant were discarded. The pellet was air dry for 10 min and resuspended in 50 μL Ultrapure water. Samples were incubated in a heat block at 65°C for 10 min and store at −20°C. Approximately 50 μg total RNA were obtained for each sample.

DNA digestion and quality control of INPUT samples

All RNA samples were digested by 2.5 μL DNase I (NEB, M0303) in 60 μL system at 37°C for 30 min. The products were purified by RNA Clean & Concentrator kit (Zymo, R1018, size limits are from 17 nt to ~23 kb). Concentration of purified RNA of all samples was diluted to 250 ng/μL by Ultrapure water before enrichment. The adjusted RNA samples were called INPUT samples. RNA integrity of each INPUT samples was detected by Fragment Analyzer (Agilent). Only INPUT samples of RQN > 8.0 were used to downstream applications.

Affinity purification of biotinylated RNA

Beads preparation

15 μL well mixed Dynabeads MyOne Streptavidin C1 (Invitrogen, 65002) was used for RNA enrichment of each sample with 25 μg RNA. Beads were resuspended with 200 μL Bead Wash Buffer (5 mM Tris-HCl, pH=7.5, 1 M NaCl (Ambion, AM9759, RNase free), 0.5 mM EDTA) and collected by a magnet for 1 min following discarding supernatant. Then this washing step was repeated twice more. The beads were washed twice with 200 μL Solution A (0.1 M NaOH, 0.05 M NaCl) with 2 min standing for RNase removing. The beads were then washed once with 200 μL Solution B (0.1 M NaCl) and blocked with 200 μL Blocking Buffer (1 mg/mL yeast tRNA, 1 mg/mL BSA, 40 ng/μL glycogen (Fermentas, R0551, RNA grade) in Ultrapure water) at room temperature for 2 hr or at 4°C overnight with thorough rotation.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

Beads binding

Blocked beads were washed 3 times with 200 μL 4 M Wash Buffer (100 mM Tris-HCl, pH7.5, 4 M NaCl, 10 mM EDTA, 0.2% Tween-20) and once with 200 μL Bead Binding Buffer (100 mM Tris-HCl, pH7.5, 1 M NaCl, 10 mM EDTA, 0.2% Tween-20). Then, beads were resuspended by 100 μL 2× Bead Binding Buffer and incubated with 100 μL INPUT at room temperature for 45 min on a rotating mixer (ThermoFisher).

Beads elution

After binding, beads were washed with 200 μL 4 M Wash Buffer for 3 times at room temperature and twice at 50°C to strip away nonspecific adsorption. Next, beads were washed twice by 200 μL RNase-free PBS (Life, AM9624) at room temperature. Discarded supernatant completely, beads were resuspended with 50 μL Elution Buffer (95% formamide, 10 mM EDTA, 1.5 mM D-biotin). The slurry was heated at 65°C for 5 min then 90°C for 5 min on a rotating mixer to achieve a good elution effect. The supernatant was collected and purified by 1 mL TRIzol Reagent for each sample. After chloroform was mixed well with sample-TRIzol mixture, RNA was dissolved into aqueous phase. The upper aqueous phase was pipetted out to a clean tube and added 20 μg glycogen to help precipitate before performing the Isopropanol precipitation. The enriched RNA was dissolved by 20 μL Ultrapure water and called ENRICH sample. 5 μL ENRICH sample was performed qRT-PCR and 5 μL ENRICH sample was used to construct a library.

All RNA-related experiments were processed in RNA workstation with RNase-free preconditioning by RNase Decontamination Wipes (Ambion, AM9788) and handled by RNase-free tubes and tips.

qRT-PCR analysis

1 μL INPUT and 5 μL ENRICH were mixed with random primers and reverse transcribed with SuperScript III (ThermoFisher, 18080044) in 20 μL total volume. Then 0.75 μL cDNA of INPUT or ENRICH was loaded to each well of a 96-well qPCR plate and performed qRT-PCR immediately with PowerUp SYBR Green Master Mix (Life, A25742) in 10 μL system by ABI StepOne Plus system. For each detected gene, there were 3 or 4 replicates for each sample (Data S6). After qRT-PCR of 40 cycles, melt curve was processed to confirm the products were unique. C_t value was averaged by all replicates. Negative controls omitting the BP probe or the unnatural nucleoside were treated in the same way as labeled samples. The enrichment fold change was calculated as $2^{\Delta C_{t\_control} - \Delta C_{t\_label}}$ and the recovery rate was calculated as $2^{-\Delta C_t}/25$, where $\Delta C_t = C_{t\_ENRICH} - C_{t\_INPUT}$.

Time course assay

HEK293T cells expressing MITO-APEX2 or APEX2-ERM were cultured as previously described. At about 60% confluence, cells were incubated with 100 μM s^6G or 100 μM s^4U for 0.5 hr, 1 hr, 2 hr, 4 hr or left untreated (for traditional APEX labeling). Following 0.5 mM BP probe incubation for 0.5 hr, cells were labeled by 1 mM H_2O_2 for 1 min at room temperature (Figures 6A–6C). Next, total RNA extraction from cells, residual DNA digestion and biotinylated RNA enrichment were carried out as described above. Then, reverse transcription and qRT-PCR were performed for INPUT and ENRICH samples. For labeling in mitochondrion, C_t values of mitochondrial RNAs (*MTCO2*, *MTCYB* and *MTND2*) and nuclear RNAs (*GAPDH*) were detected by qRT-PCR. For labeling at ERM, C_t values of secretory RNAs (*SSR2*, *TMX1* and *SFT2D2*) and non-secretory RNAs (*FAU*, *SUB1* and *MTCO2*) were detected by qRT-PCR. C_t values were normalized by INPUT samples for each RNA.

Transcriptional inhibition assay

HEK293T cells expressing APEX2-LMNB1 were cultured as described above. At approximately 60% confluence, cells were treated with 2 μg/mL actinomycin D (ActD) for 4 hr, 6 hr, or left untreated (Schofield et al., 2018). For MERR APEX labeling, cells were incubated with 100 μM s^6G for 5 hr. For traditional APEX labeling, no additional manipulations for cells. Following 0.5 mM BP probe incubation for 30 min, cells were labeled by 1 mM H_2O_2 for 1 min (Figures 6D–6F). Total RNA extraction, DNA digestion and affinity purification of biotinylated RNA steps were performed as previously described. Reverse transcription and qRT-PCR were carried out to detect RNA in nucleus (*HOXD10*, *KIAA0907*, *XIST*, *SUB1* and *GAPDH*) and mitochondrion (*MTCO2*). C_t values were normalized by INPUT samples for each RNA.

Next generation sequencing

1 μL INPUT and 5 μL ENRICH (corresponding to approximately 6-7 μg unenriched INPUT) was used for library construction by NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB, E7770) with Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) following the instructions. In the process of library construction, NEBNext Sample Purification Beads were replaced by VAHTS DNA Clean Beads (Vazyme, N411) which were needed to transfer from 4°C to room temperature for at least 30 min before using. After isolation, mRNA was fragmented to approximately 300 nt at 94°C for 12 min. First Strand Synthesis was then performed with random primers. After Second Strand Synthesis, double-strand cDNA was purified by 1.8× DNA Clean Beads. Purified cDNA libraries were assembled by End Prep Reaction following Adaptor Ligation. The ligation mixture was purified by 0.9× DNA Clean Beads to remove redundant adaptors. Purified cDNA was proceeded to PCR experiments with 11 (INPUT samples) or 15 (ENRICH samples) cycles. 0.9× DNA Clean Beads were then used to remove extra primers and adaptors dimers. Quality of Libraries was controlled by Fragment Analyzer. If primers or adaptor dimers remained, a further size selection with 0.9× DNA Clean Beads needed to be performed. If the peak got a wide distribution, two rounds of size selection was needed with 0.8× and 0.2× DNA Clean Beads according to both NEBNext Ultra II RNA Library Prep Kit for Illumina and VAHTS DNA Clean Beads manuals. All good-quality cDNA libraries were sequenced of 150 bp paired-end approximately 10 M reads on the Illumina HiSeq X Ten platform.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

Single-molecule FISH imaging

Primary probes for targeted mRNAs were designed by Oligostan (Tsanov et al., 2016) and have been published in CAP-seq (Wang et al., 2019). Primary probes were dissolved at final concentration of 100 μM by TE buffer (pH = 8.0). Mixed every equimolar primary probe of each mRNA and diluent 5 times in TE buffer at final concentration of 20 μM. Secondary probes conjugated with Alexa Fluor 488 was dissolved at 100 μM by TE buffer named FLAP. 40 pmol mixed primary probes and 50 pmol FLAP were dissolved in NEB3.1 (NEB, B7203V) buffer in 10 μL to get duplex. The duplex was incubated at 85°C for 3 min, 65°C for 3 min, 25°C for 5 min and hold at 4°C. HEK 293T cells were plated on Matrigel matrix coated coverslips in 24-well plate to be 60-70% confluent. Cells were rinsed once with RNase-free PBS and incubated with freshly made 15% formamide in 1× SSC buffer at room temperature for 15 min. Mix1 (5 μL 20× SSC, 1.7 μL 20 μg/μL yeast tRNA, 15 μL formamide, 2 μL duplex and 26.3 μL H_2O) and mix2 (1 μL 20 μg/μL BSA, 1 μL 200 mM RNase inhibitor VRC, 26.5 μL 40% dextran sulphate sodium salt and 21.5 μL H_2O) were placed on ice and mix1 was added to mix2. 50 μL of the hybridization mix for each sample was dropped in a 10 cm dish and the coverslips with cells were covered on the drop (cells facing down). A 3.5 cm dish containing 1 mL of 15% formamide in 1× SSC solution was put inside the 10 cm dish. The 10 cm dish was sealed by parafilm and incubated at 37°C overnight. The coverslips were placed inside a 24-well plate and washed twice by freshly prepared 15% formamide in 1× SSC at 37°C for 30 min, and then rinsed twice with PBS. Cells on the coverslips were permeabilized again with 0.1% Triton X-100 in RNase-free PBS. Then the coverslips were blocked in 3% BSA in RNase-free PBS with 2 mM RVC for 1 hr at room temperature. 1:1000 rabbit-anti-laminA in 3% BSA in RNase-free PBS was incubated with 2 mM RVC for 1 hr at room temperature. The coverslips were washes 3 times by RNase-free PBS with 2 mM RVC for 5 min. 1:1000 anti-rabbit-647 and 1:1000 DAPI were incubated in 3% BSA in RNase-free PBS with 2 mM VRC for 1 hr. After rinsed twice with RNase-free PBS, the cells were imaged on an inverted fluorescence microscope platform with confocal unit the same as immunofluorescence imaging. Images were analyzed by ImageJ (v2.1.0/1.53d) (Schindelin et al., 2012; Schneider et al., 2012).

For statistical analysis of single molecule FISH imaging, the confocal image stack was maximum intensity-projected along the z-axis. For each RNA, the intensities of DAPI channel were used to identify the nucleus and the cell membrane boundary (aided by the weak and non-specific cellular adsorption background of DAPI). The FISH channel was converted into binary masks using manually chosen threshold, to identify individual RNA puncta. The ratio of nucleus-localized puncta over total number of puncta was then calculated for each image stack. To evaluate the influence of s^6G metabolism experiments, Kruskal-Wallis test was used to compare the puncta ratio in the presence versus absence of s^6G.

QUANTIFICATION AND STATISTICAL ANALYSIS

All graphics were generated by R (the most commonly used package was ggplot2 (v3.3.2) (Wickham, 2016)), GraphPad Prism 9 or Microsoft Excel 2019 for Mac (v16.40).

NGS data analysis

The RNA-seq reads were quality controlled by FastQC (v0.11.8) then summarized by MultiQC (v1.8) (Ewels et al., 2016). After quality control, the sequencing reads were mapped using hisat2 (v2.1.0) (Kim et al., 2015) to the human genome assembly GRCh38 (hg38) with gene annotation (v.87) downloaded from Ensembl website (Yates et al., 2020). The mapped reads were counted by htseq-count (v0.7.2) (Anders et al., 2015) with the option '–stranded no'.

For mitochondrial matrix enrichment samples, we calculated TPM (Transcripts Per Kilobase Million) to distinguish the highest differential expression genes. Only genes with TPM > 0 were considered in this part.

For NGS data of endoplasmic reticulum membrane (ERM), differential expression analysis was carried out by R package DESeq2 with MERR APEX-seq raw data of ERM versus NES to obtain $\log_2$ fold change. To improve the quality of analysis, we filtered genes at low expression level (DESeq2 baseMean > 100). The receiver-operator-curve (ROC) was generated to test the sensitivity and specificity of dataset at ERM. The true positive list was ER-enriched transcripts using ribosome profiling (Jan et al., 2014) and the false positive RNAs were gathered by transcripts not in Phobius (Kall et al., 2004), SignalP (Petersen et al., 2011) and TMHMM (Krogh et al., 2001). From the ROC analysis, a cutoff of $\log_2$ fold change (ERM versus NES) > 0.27 was determined to optimize specificity (0.94) and sensitivity (0.85). In addition, DESeq2 results was filtered by P value less than 0.05 to generate confident ERM dataset. The final list of enriched genes is summarized in Data S2.

GOCC-secretome was generated according to the GOCC annotations related to the secretory pathway, including ER, smooth endoplasmic reticulum, rough endoplasmic reticulum, endoplasmic reticulum-Golgi intermediate compartment, Golgi apparatus, Golgi cis cisterna, cis-Golgi network, trans-Golgi network, vesicles, peroxisomal membrane, trans-Golgi network transport vesicle membrane, transport vesicle membrane, cytoplasmic vesicle membrane, plasma membrane and extracellular region. HPA-secretome was defined from HPA database with annotations containing 'Endoplasmic Reticulum', 'Golgi Apparatus', 'Plasma Membrane', 'Secreted proteins' and 'Vesicles'. List of predicted transcripts at ERM is a collection of GOCC-secretome and HPA-secretome.

For LMNB1 NGS data, For LMNB1 NGS data, differential expression genes were analyzed by R package DESeq2 (v1.29.6) (Gentleman et al., 2004; Love et al., 2014). Raw counts were normalized to eliminate the impact of sequencing depth. The cutoff was set to log2 (ENRICH versus INPUT) > 0.65 and FDR < 0.05.

For LADs analysis, we gathered Gene stable ID from reported LADs genomic regions (Guelen et al., 2008). Gene annotation file of hg18 is the UCSC Table Browser data retrieval tool (Karolchik et al., 2004). The annotation of LADs was performed by bedtools intersect (v2.27.1) (Quinlan and Hall, 2010) then converted into Gene stable ID.

Please cite this article in press as: Li et al., Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome, Cell Chemical Biology (2022), https://doi.org/10.1016/j.chembiol.2022.02.005

NGS data visualization

For mitochondrial RNA enrichment data visualization, reads were then generated from BAM file to BigWig file using bamCoverage (v2.5.3) (Ramirez et al., 2014) with option '–scaleFactor' to normalize all mitochondrial RNA counts to 2,000,000 of each sample. Then BigWig files of two biological replicates were merged into one bedGraph file using bigWigMerge (v2) (Kent et al., 2010). BedGraph files were converted to BigWig files using bedGraphToBigWig (v4) (Kent et al., 2010) with chromosome size extracted from GRCh38. For LMNB1 MERR APEX-seq and APEX-seq data visualization, the '–scaleFactor' was normalized by all mitochondrial RNA counts to 200,000 of each sample. The visualization of all data was performed in Integrative Genomics Viewer (Robinson et al., 2011, 2017; Thorvaldsdottir et al., 2013) and WashU Epigenome Browser (Zhou et al., 2011, 2015).

Statistical methods

To identify mutations in MITO APEX-seq data, we performed SNP calling on the reads aligned uniquely (flag: 83/163, 99/147) with MAPQ>2 of INPUT and ENRICH samples. SNPs were detected using SAMtools (Li et al., 2009) mpileup programs over hg38 genome following sorted reads combination of the replicates. Confident SNPs were filtered based on the following criteria: 1) QUAL values greater than 20; 2) raw read depth larger than 30.

For all sequencing data, we did correlation analysis and T-distributed Stochastic Neighbour Embedding (t-SNE) test (van der Maaten and Hinton, 2008). Raw counts were melted by R package reshape2 (v1.4.4) (Zhang, 2016) and reordered to get correlation plot. And raw counts of all samples greater than zero were analyzed by R package Rtsne (v0.15) (Krijthe, 2015) to get t-SNE plot.

Gene Ontology Cellular Component (GOCC) enrichment analysis for nuclear lamina enrichment experiments was identified as the ensemble of genes whose GOCC terms annotations include subcellular locations related to nuclear components with Fisher's exact test (Ashburner et al., 2000; Carbon et al., 2009; Day-Richter et al., 2007; Mi et al., 2019; The Gene Ontology, 2019).

Gene Ontology Biological Process (GOBP) enrichment analysis was performed in R with packages org.Hs.eg.db (3.11.4) (Carlson, 2020), clusterProfiler (v3.17.0) (Yu et al., 2012), stringr (v1.4.0) (Wickham, 2019) and Rgraphviz (v2.33.0) (Hansen et al., 2020).

Splicing events distributions were quantified by rMATs (Shen et al., 2014). The bam files of ENRICH samples against INPUT samples were performed under arguments -t paired, with gtf gene annotation file (v.87) of hg38 downloading from Ensembl website. The statistic model of rMATS has been normalized by the lengths of individual splicing variants. Splicing difference was filtered by FDR < 0.05 in output files.