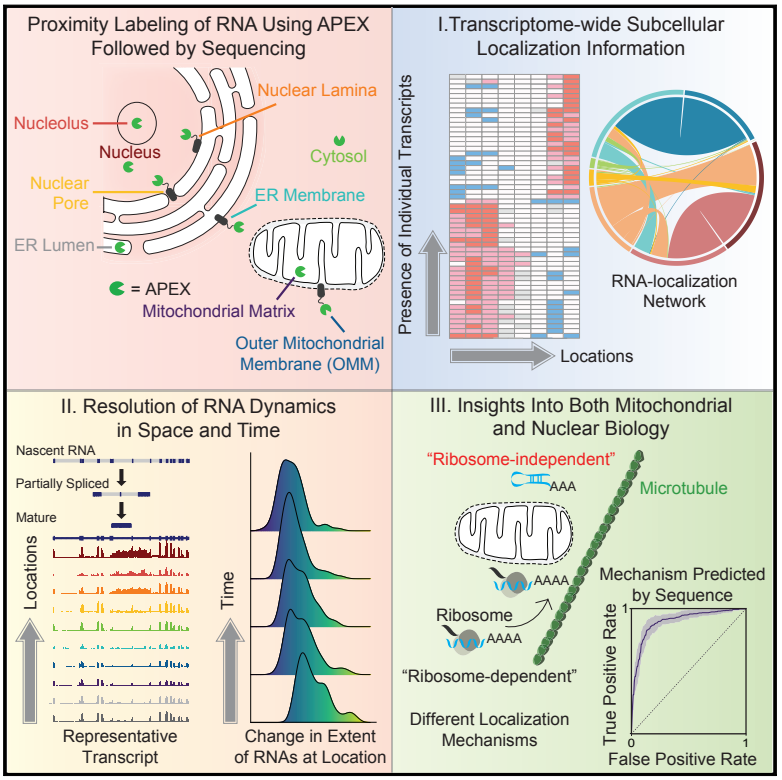
Resource

Cell

Atlas of Subcellular RNA Localization Revealed by APEX-Seq

Graphical Abstract



Highlights

- A transcriptome-wide subcellular RNA atlas was generated by proximity labeling
- Isoform-level subcellular localization patterns for over 3,200 genes identified
- RNA-transcript location correlates with genome architecture and protein localization
- Two modes of mRNA localization to the outer mitochondrial membrane uncovered

Authors

Furqan M. Fazal, Shuo Han, Kevin R. Parker, ..., Alistair N. Boettiger, Howard Y. Chang, Alice Y. Ting

Correspondence

howchang@stanford.edu (H.Y.C.), ayting@stanford.edu (A.Y.T.)

In Brief

A newly developed technique reveals the subcellular transcriptomes at many landmarks in the nucleus and cytosol and connects mRNA localization to genome architecture, protein location, and local-translation mechanisms.

Fazal et al., 2019, Cell *178*, 473–490
July 11, 2019
https://doi.org/10.1016/j.cell.2019.05.027

CellPress

Resource

Cell

Atlas of Subcellular RNA Localization Revealed by APEX-Seq

Furqan M. Fazal,[1,2,3,9] Shuo Han,[3,4,5,9] Kevin R. Parker,[1,2,3,10] Pornchai Kaewsapsak,[3,4,5,10] Jin Xu,[1,2,3] Alistair N. Boettiger,[6] Howard Y. Chang,[1,2,3,7,*] and Alice Y. Ting[3,4,5,8,11,*]

[1]Center for Personal Dynamics Regulomes, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Department of Dermatology, Stanford University School of Medicine, Stanford, CA 94305, USA
[3]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Department of Chemistry, Stanford University, Stanford, CA 94305, USA
[5]Department of Biology, Stanford University, Stanford, CA 94305, USA
[6]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[7]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA
[8]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[9]These authors contributed equally
[10]These authors contributed equally
[11]Lead Contact
*Correspondence: howchang@stanford.edu (H.Y.C.), ayting@stanford.edu (A.Y.T.)
https://doi.org/10.1016/j.cell.2019.05.027

SUMMARY

We introduce APEX-seq, a method for RNA sequencing based on direct proximity labeling of RNA using the peroxidase enzyme APEX2. APEX-seq in nine distinct subcellular locales produced a nanometer-resolution spatial map of the human transcriptome as a resource, revealing extensive patterns of localization for diverse RNA classes and transcript isoforms. We uncover a radial organization of the nuclear transcriptome, which is gated at the inner surface of the nuclear pore for cytoplasmic export of processed transcripts. We identify two distinct pathways of messenger RNA localization to mitochondria, each associated with specific sets of transcripts for building complementary macromolecular machines within the organelle. APEX-seq should be widely applicable to many systems, enabling comprehensive investigations of the spatial transcriptome.

INTRODUCTION

The subcellular localization of RNA is intimately tied to its function (Buxbaum et al., 2015). Asymmetrically distributed RNAs underlie organismal development, local protein translation, and the 3D organization of chromatin. Where an RNA is located within the cell likely determines whether it will be stored, processed, translated (Berkovits and Mayr, 2015), or degraded (Fasken and Corbett, 2009).

While many methods have been developed to study RNA localization (Weil et al., 2010), only a few have been applied on a transcriptome-wide scale. The most classic approach is biochemical fractionation to enrich specific organelles, followed by RNA sequencing ("fractionation-seq"). However, a major limitation of fractionation-seq is that it cannot be applied to organelles that are impossible to purify, such as the nuclear lamina and outer mitochondrial membrane (OMM). Even for organelles that *can* be enriched by centrifugation, such as mitochondria, current protocols fail to remove contaminants (Sadowski et al., 2008).

RNA localization can also be directly visualized by microscopy (Bertrand et al., 1998; Femino et al., 1998), and techniques have recently been pioneered for imaging thousands of cellular RNAs at once using barcoded oligonucleotides (Chen et al., 2015b; Shah et al., 2016). The drawbacks of these fluorescence *in situ* hybridization (FISH)-based approaches, however, are the need for designed probe sets targeting RNAs of interest; the requirement for cell fixation and permeabilization, which can relocalize cellular components (Fox et al., 1985; Schnell et al., 2012); the difficulty of assigning RNAs to specific cellular landmarks due to spatial resolution limits; and the limited information content compared to RNA sequencing. Finally, these transcriptome-wide imaging methods are technically challenging and require specialized instrumentation not available to most.

An adaptation of ribosome profiling (Ingolia et al., 2009) has enabled this technique to profile actively translated mRNAs in specific cellular locales. The two demonstrations—on the endoplasmic reticulum membrane (ERM) in yeast and mammalian cells (Jan et al., 2014), and on the OMM in yeast (Williams et al., 2014)—showed high spatial specificity and compatibility with living cells. However, the methodology cannot detect non-coding RNAs or non-translated mRNAs. Proximity-specific ribosome profiling is also not yet a fully generalizable method, as the requirement for biotin starvation during cell culture is prohibitively toxic to many cell types.

Hence, there remains a need for new methodology that can map the spatial localization of thousands of endogenous RNAs at once in living cells. The method should be applicable to any subcellular region and capture full sequence details of any RNA type, enabling comparisons across RNA variants and isoforms. Here, we develop the "APEX-seq" methodology in an effort to provide these capabilities. We characterize the

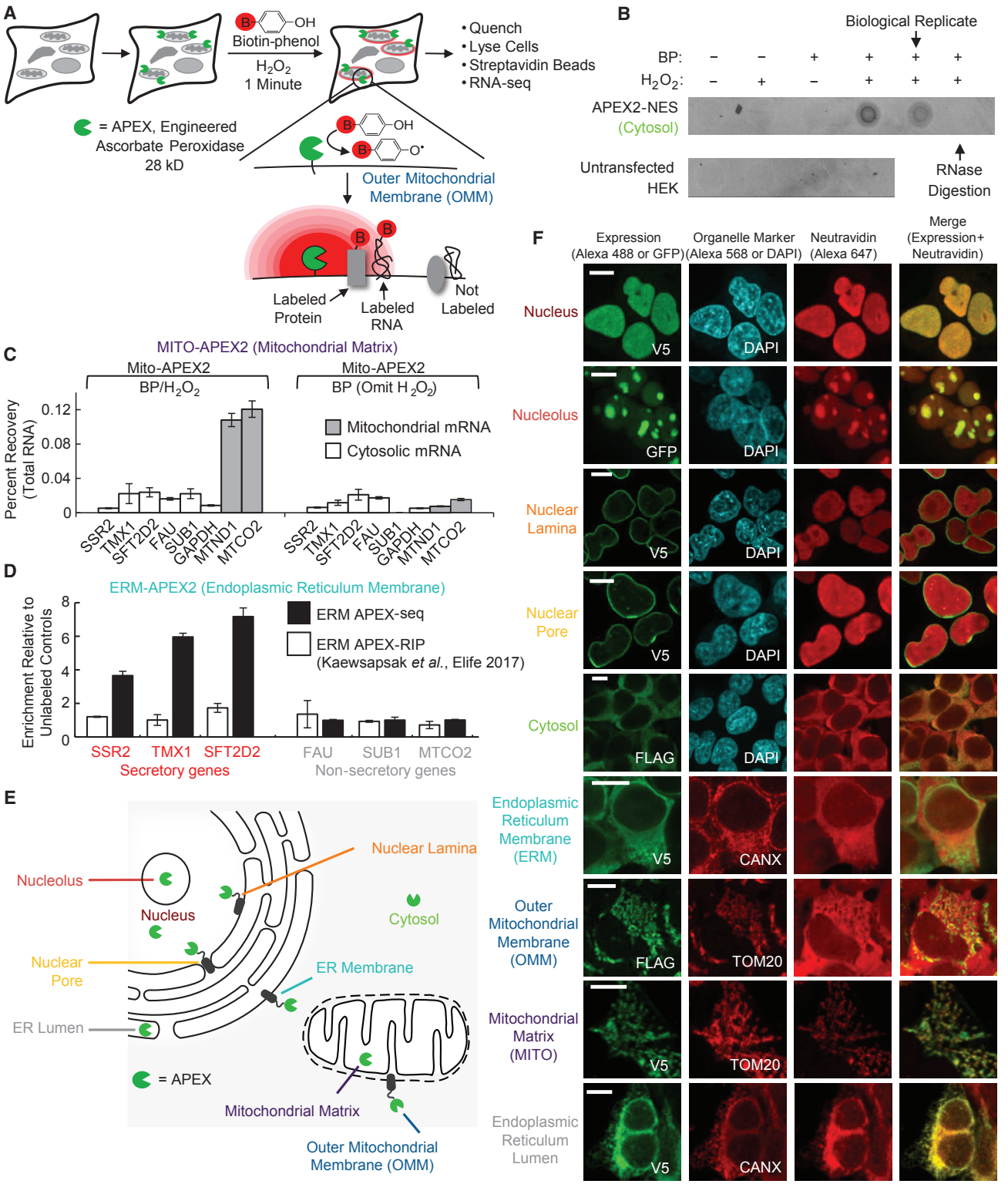


Figure 1. Development of APEX-Seq Methodology

(A) APEX2-mediated proximity biotinylation of endogenous RNAs. APEX2 peroxidase is genetically targeted to the cellular region of interest. Addition of BP (red B = biotin) and H_2O_2 to live cells for 1 min results in biotinylation of endogenous proteins and RNA within a few nanometers of APEX2. Biotinylated RNAs are separated using streptavidin-coated beads, poly(A)-selected, and analyzed by RNA sequencing (RNA-seq).

(legend continued on next page)

APEX-seq approach and then apply it to nine subcellular locations, generating a high-resolution atlas of endogenous RNA localization in living human HEK293T cells. Our data reveal correlations between localization of mRNAs and the protein products they encode, as well as patterns of RNA localization and underlying genome architecture. An analysis of mRNAs at the OMM suggests distinct mechanisms for RNA targeting that correlate with the sequence and function of the encoded mitochondrial proteins. These examples illustrate the versatility of APEX-seq and its ability to nominate or test novel biological hypotheses.

RESULTS

APEX-Catalyzed Labeling of RNA

To develop the methodology, we drew from previous work in our laboratory using enzymes to map spatial proteomes (Rhee et al., 2013). APEX2 (Lam et al., 2015) is an evolved mutant of soybean ascorbate peroxidase that catalyzes the one-electron oxidation of biotin-phenol (BP), a membrane-permeable small molecule. The resulting BP radical is short-lived (half-life <1 ms) (Mortensen and Skibsted, 1997; Wishart and Rao, 2010) and covalently conjugates onto protein side chains. Hence, APEX2 catalyzes the promiscuous biotin tagging of endogenous proteins within a few nanometers of its active site in living cells. The high spatial specificity of this approach has enabled APEX mapping of numerous organelle proteomes as well as protein interaction networks (Han et al., 2018).

We previously combined APEX proteomic tagging with formaldehyde protein-RNA crosslinking in order to extend our analysis to cellular RNAs (Kaewsapsak et al., 2017). While this "APEX-RIP" approach was effective at mapping the RNA composition of membrane-enclosed organelles such as the mitochondrion, its spatial specificity was poor in "open" or non-membrane enclosed cellular regions. For instance, RNAs enriched by APEX targeted to the ERM (facing cytosol) were no different from those enriched by cytosolic APEX. A version of this two-step strategy using UV crosslinking may improve specificity (Benhalevy et al., 2018).

A more straightforward and potentially higher-specificity approach would be to bypass crosslinking altogether and use APEX peroxidase to directly biotinylate cellular RNAs within a short time window (Figure 1A). To test whether peroxidase-generated phenoxyl radicals could biotinylate RNA *in vitro*, we combined horseradish peroxidase (HRP), which catalyzes the same one-electron oxidation chemistry as APEX2, with tRNA, BP, and H_2O_2. On a streptavidin dot blot, we observed robust tRNA biotinylation that was abolished by RNase treatment but unaffected by proteinase K treatment (Figure S1A). We next used a RT stop assay to evaluate the labeling and found that, while full-length transcripts are still produced, multiple RT stops are observed at G-rich regions in peroxidase-catalyzed RNA samples (Figures S1D and S1E). Additional experiments characterized the covalent adduct between G and BP by HPLC and mass spectrometry (Figures S1B and S1C).

To test APEX-catalyzed RNA biotinylation in living cells, we generated HEK cells stably expressing APEX2 in the cytosol. We labeled the cells with BP and H_2O_2 for 1 min, extracted total RNA, and analyzed the RNA by streptavidin dot blot. Figure 1B shows that RNA biotinylation is abolished upon omission of BP or H_2O_2 or following treatment with RNase. Combined with the assays above, our results suggest that APEX directly tags RNA with biotin, not merely biotinylating proteins co-complexed with RNA.

Next, we combined APEX labeling with qRT-PCR analysis of biotinylated RNAs in order to begin assessing the spatial specificity of this approach. We started with the mitochondrial matrix, which we have previously characterized by APEX proteomics (Han et al., 2017; Rhee et al., 2013), and whose transcriptome can be predicted by the sequence of the mitochondrial genome (mtDNA) (Mercer et al., 2011). Using HEK cells expressing APEX2 in the mitochondrial matrix, we performed labeling and then extracted RNA and enriched the biotinylated fraction using streptavidin beads. We optimized a series of denaturing washes to fully dissociate complexes and ensure that the streptavidin beads only enriched biotinylated RNA species (Figure S1F). We then analyzed the eluate by qRT-PCR and observed strong enrichment of mtDNA-encoded mRNAs *MTND1* and *MTCO2* but not negative-control cytosolic mRNAs (Figure 1C).

However, because the mitochondrial matrix is enclosed by a tight membrane that is impervious to BP radicals (Rhee et al., 2013), it does not provide a rigorous test of APEX labeling radius. To evaluate spatial specificity in an open cellular compartment, we utilized HEK cells stably expressing APEX2 on the ERM, facing cytosol. qRT-PCR analysis of streptavidin-enriched RNA following BP labeling (Figure 1D) shows high enrichment of secretory mRNAs (ERM-proximal "true positives") but not negative-control cytosolic mRNAs (encoding non-secretory proteins). This result suggests that APEX biotinylation has nanometer spatial resolution and is able to distinguish ER-proximal RNAs

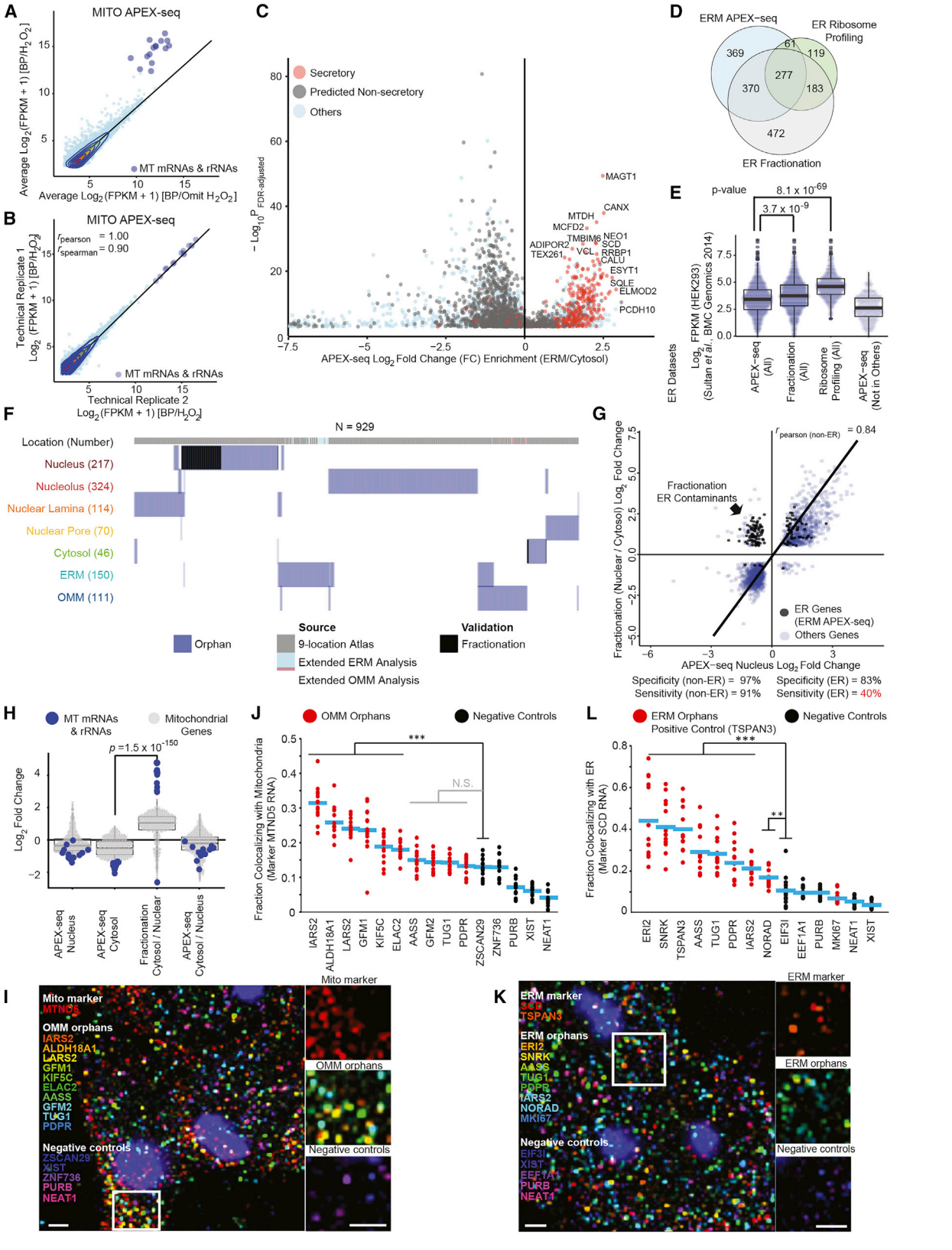
(B) Streptavidin-biotin dot-blot analysis of direct RNA biotinylation by APEX2 in cells. HEK-293T cells expressing APEX2 in the cytosol were labeled with for 1 min and then the RNA was extracted and blotted. Only when BP, H_2O_2, and APEX2 were all present was the signal observed. RNase treatment of the sample abolished the signal.

(C) qRT-PCR analysis showing specific enrichment of mitochondrial RNAs (gray) over cytosolic mRNAs (white). Cells expressing APEX2 targeted to the mitochondrial matrix were labeled for 1 min. Biotinylated RNAs were enriched following RNA extraction. Data are the mean of 4 replicates ± 1 SD.

(D) qRT-PCR analysis showing specific enrichment of secretory (red) over non-secretory (gray) mRNAs with APEX-seq, but not APEX-RIP. Cells stably expressing APEX2 targeted to the ERM membrane (facing cytosol) were labeled for 1 min. For APEX-RIP, RNAs were crosslinked to proteins for 10 min before streptavidin beads enrichment. Data are the mean of 4 replicates ±1 SD. The data were normalized such that the mean enrichment of non-secretory RNAs was 1 for both techniques.

(E) Human cell showing nine different subcellular locations investigated.

(F) Fluorescence imaging of APEX2 localization and biotinylation activity. Live-cell biotinylation was performed for 1 min in cells stably expressing the indicated APEX2 fusion protein. APEX2 expression was visualized by GFP or antibody staining (green). Biotinylation was visualized by staining with neutravidin-Alexa Fluor 647 (red). DAPI is a nuclear marker. Endogenous *TOM20* and *CANX* were used as markers for the mitochondria and ER, respectively. Scale bars, 10 μm.
See also Figure S1 and Table S1.



(legend on next page)

from cytosolic RNAs only nanometers from the ERM. This result strikingly contrasts with previous observations using APEX-RIP (Kaewsapsak et al., 2017). For a further side-by-side comparison between APEX-seq and APEX-RIP, a total of 8 representative transcripts that are known to localize to the respective landmarks based on previous literature, were investigated by qRT-PCR (Figure S2E). APEX-seq enriched specific, proximal RNAs in open subcellular regions (ERM, nuclear lamina, nucleolus, and OMM), whereas APEX-RIP was unable to do so.

Development and Validation of APEX-Seq

Encouraged by the results above, we moved to a more comprehensive analysis by replacing qRT-PCR with transcriptome-wide sequencing. We also created cell lines expressing APEX in nine subcellular locales (Figures 1E and S2A). For each cell line, we verified correct targeting of APEX by performing immunofluorescence staining against organelle markers. To examine APEX activity, we performed BP labeling, fixed, and stained the biotinylated species using neutravidin-Alexa 647. For some locations, the neutravidin pattern overlapped closely with APEX localization (e.g., nucleolus and mitochondrial matrix; Figure 1F), indicating minimal diffusion of biotinylated species. For other locations, the neutravidin signal was more "spread out" than the APEX signal (e.g., ERM and OMM; Figure S2B), suggesting redistribution of biotinylated species during the 1-min labeling time window (Hung et al., 2016).

To assess the quality of the poly(A)-selected APEX-seq data (Figures S2C and S2D; Table S2), we first focused on two subcellular compartments that have been extensively mapped: the mitochondrial matrix and the ERM. For the former (Figures 2A and 2B), APEX-seq experiments showed strong enrichment of all 13 mRNAs and the 2 rRNAs encoded by mtDNA (Figures S2F and S2G), while no RNAs encoded by the nuclear genome were highly enriched.

For the ERM, APEX-seq highly enriched RNAs previously shown to be ER proximal (such as mRNAs encoding secreted proteins) over cytosol-localized RNAs. To perform a quantitative analysis, we used ROC cutoff analysis (Linden, 2006) (Figures S2I and S2J) to produce a list of 1,077 ERM-enriched RNAs (Figure 2C). To evaluate the specificity of this dataset, we determined the fraction of "secretory" or "transmembrane" mRNAs (STAR Methods) and found that 90% of genes had such prior annotations. The remaining 10% (107 genes) could be false-positives, or they could be newly discovered ERM-associated RNAs.

To evaluate depth of coverage, we prepared a hand-curated list of 71 well-established ER-resident proteins and asked what fraction of their corresponding mRNAs appear in our ERM APEX-seq dataset. We recovered 70% of this true-positive list (Figure S2K). This sensitivity is comparable to that of our previous APEX proteomic datasets in open compartments (Hung et al., 2017) (Figure S2L). RNAs we failed to enrich could be sterically shielded in the live cell environment, low in abundance, or dual-localized to both ERM and cytosol.

The ERM-associated transcriptome has previously been studied by fractionation-seq (Reid and Nicchitta, 2012) and proximity-specific ribosome profiling (Jan et al., 2014). Upon analyzing the published datasets, we found that the specificities of fractionation-seq and APEX-seq were comparably high (90% versus 91% secretory mRNAs, respectively; Figure S2I), in addition to sensitivity (Figure S2K). However, Figure 2D shows that each method recovers somewhat different subsets of transcripts. Further analysis of genes enriched by APEX-seq but *not* fractionation-seq or ribosome profiling show that many of these are lower in RNA abundance (Figure 2E).

Altogether, our APEX-seq analysis demonstrates that high specificity and reasonable sensitivity can be achieved in both membrane-enclosed and open subcellular compartments.

RNA Atlas of 9 Distinct Subcellular Compartments by APEX-Seq

Having established the specificity and sensitivity of APEX-seq using the mitochondrial matrix and ERM, we turned our attention to the seven other compartments (Figure 1E). The RNA content of most of these regions has not previously been mapped, as

Figure 2. Validation of APEX-Seq, Including Specific Orphans from RNA Atlas

(A) APEX-seq in the mitochondrial matrix. Transcript abundance in experiment plotted against negative control (omit H_2O_2). All mRNAs and rRNAs encoded by the mitochondrial genome (large blue dots) are enriched by APEX (mean enrichment >11-fold). FPKM, fragments per kilobase of transcript per million reads. Due to the 100-nt size selection step during RNA extraction, tRNAs were not efficiently recovered.

(B) Scatterplot of transcript abundance in the mitochondrial matrix (MITO).

(C) APEX-seq at the ERM, facing cytosol. Volcano plot showing APEX-catalyzed enrichment of secretory mRNAs (red) over non-secretory mRNAs (black).

(D) Comparison of ERM-enriched RNAs by APEX-seq, proximity-specific ribosome profiling, and ER fractionation-seq.

(E) Transcript abundance (FPKM) analysis of genes enriched by ERM APEX-seq, fractionation-seq, proximity-specific ribosome profiling, and genes unique to the APEX-seq dataset. p values are from a Mann-Whitney U test.

(F) Total number of orphans (blue) generated from APEX-seq RNA datasets, with those validated by further poly(A)+ fractionation-seq shown in black. The source of most of these RNAs is the RNA atlas, with further contributions from analysis of the ERM and OMM transcriptomes.

(G) APEX-seq yields cleaner results than bulk fractionation RNA-seq. Nucleus APEX-seq fold changes are highly correlated with bulk fractionation RNA-seq when considering non-ER genes (blue). However, fractionation suffers from contamination by ER transcripts (black).

(H) APEX-seq in the cytosol does not recover mitochondrial-genome encoded RNAs, whereas fractionation-seq does. mRNAs and rRNAs encoded by the mitochondrial genome are shown in blue, whereas mRNAs for mitochondrial proteins encoded by the nuclear genome are shown in grey. p value is from a Mann-Whitney U test.

(I and K) Sequential smFISH imaging of OMM (I) or ERM (K) orphans in HEK cells. MTND5 was used as a mitochondrial marker. SCD and TSPAN3 were used as ERM markers. mRNAs and lncRNAs not enriched in OMM (I) or ERM (K) were used as negative controls. Expanded views of the boxed region are shown on the right. Scale bar, 5 μm.

(J and L) Quantitation of OMM (J) or ERM (L) orphans colocalization with MTND5 (J) or SCD (L) by sequential smFISH imaging. Blue lines represent mean from 14 independent fields of view. Data were analyzed using a two-tailed Student's t test, with *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$; N.S., not significant ($p > 0.05$). See also Figure S2 and Tables S2 and S3.

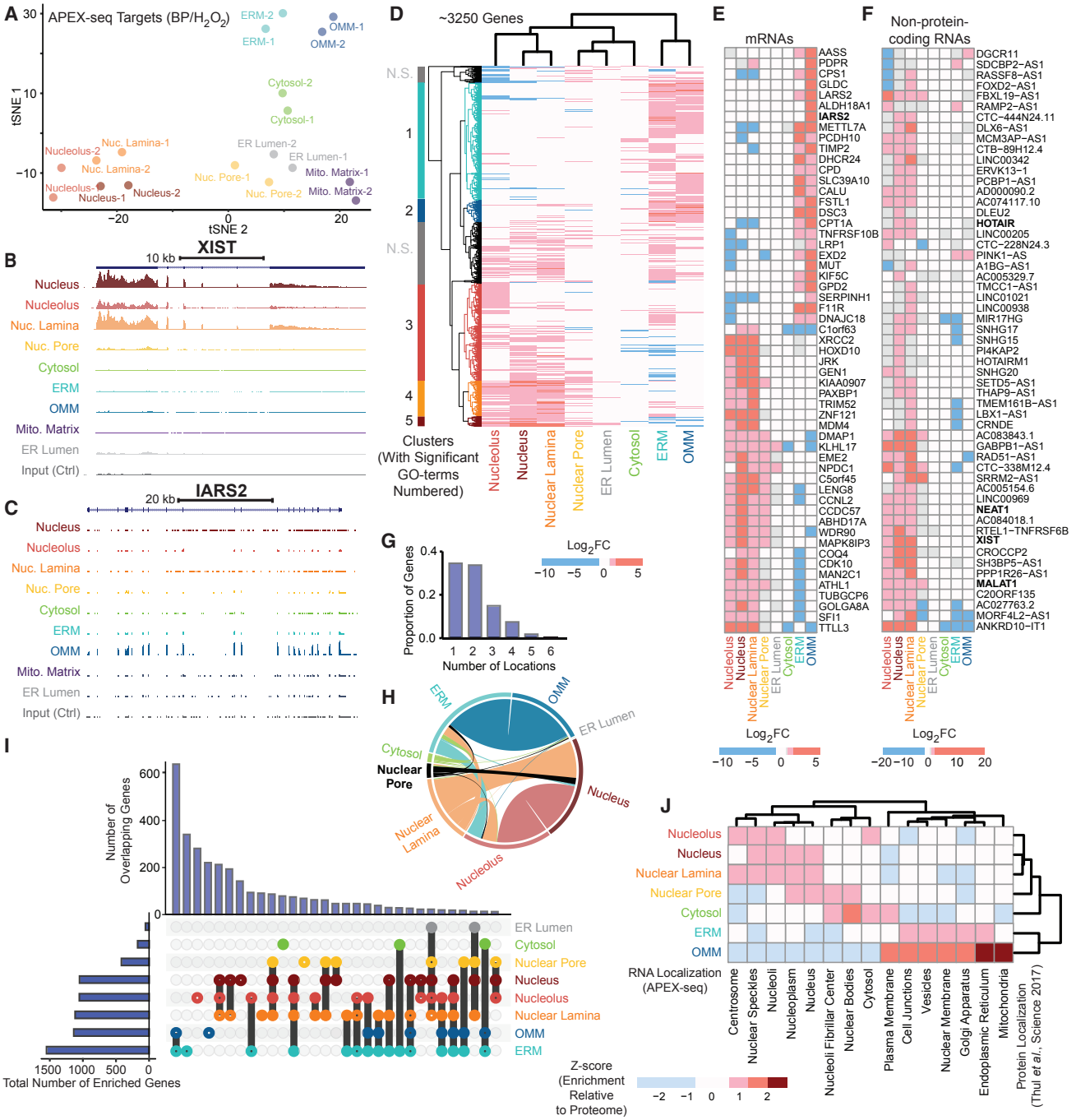


Figure 3. Analysis of Subcellular Transcriptome Maps

(A) T-distributed stochastic neighbor embedding (t-SNE) plot showing separation and clustering of APEX-seq libraries.

(B and C) Genome tracks for *XIST* (B), a nuclear non-coding RNA, and (C) *IARS2*, an mRNA encoding a mitochondrial tRNA synthetase. For each location, the reads were averaged across two APEX-seq replicates. The control tracks were generated by averaging 18 controls from all 9 constructs.

(D) Heatmap of transcripts enriched by APEX-seq showing clustering of the genes that specifically localize to at least one location and have fold-change data from all locations.

(E) Heatmap showing the APEX-seq fold changes for the mRNA transcripts found to be most variable among the locations investigated.

(F) Heatmap showing the APEX-seq fold changes for non-coding RNAs (excluding pseudogenes) that have the most-variable localization enrichment. A few well-known noncoding RNAs are shown in bold.

(G) Of the ~3,250 genes analyzed, most localize to only one or two of the eight locations (excluding mitochondrial matrix) interrogated.

(legend continued on next page)

they are impossible to purify and/or too small to image unambiguously by conventional microscopy. As such, it is impossible to generate true positive and false positive lists of known resident and non-resident RNAs respectively with which to perform ROC-based cutoff analysis. We therefore opted for a universal enrichment-factor cutoff of 0.75 (log_2 fold change) and q value (false discovery rate [FDR]-adjusted p value) cutoff of 0.05, which was applied to all compartments (STAR Methods). By intersecting data from each pair of replicates, we obtained RNA lists for all nine compartments (Table S3).

These lists provide a wealth of observations about the RNA composition of diverse cellular locales. Many RNAs are "orphans," never previously linked to the compartment to which APEX-seq assigns them. For instance, our APEX-seq atlas (Figure 2F) newly assigns 324 RNAs to the nucleolus, 114 RNAs to the lamina, and 111 RNAs to the OMM. To provide further confidence in these spatial assignments, we analyzed a subset of high-abundance RNAs by sequential smFISH imaging (Figure 2I–2L) and found that 6 out of 10 OMM orphans and 7 out the 8 ERM orphans displayed significant smFISH enrichment at the mitochondria and ERM, respectively.

To further validate nuclear and cytosolic RNAs enriched by APEX-seq, we performed poly(A)+ nuclear/cytosolic fractionation of matched HEK cells (Figure 2G). Of the 95 nuclear and 14 cytosolic APEX-seq orphans for which we could obtain high-quality fractionation-seq reads, 84 of the nuclear and 4 of the cytosolic RNAs were validated (Figure 2F). Overall, fractionation-seq validated 81% (n = 88/109) of orphan genes.

The availability of matched fractionation-seq datasets gives us the opportunity to compare head-to-head with APEX-seq. Overall, we found that both nuclear and cytosolic APEX-seq datasets were much more specific than our corresponding fractionation-seq data. For instance, our APEX-seq gene lists lacked the mitochondrial matrix and ER contaminants present in the cytosolic and nuclear fractionation data, respectively (Figures 2G, 2H, and S3F). Excluding ER transcripts in the nuclear fractionation-seq dataset (using ERM APEX-seq gene list), we compared the remaining genes to APEX-seq in order to estimate the accuracy (94%) and precision (96%) of our methodology. We also observed that the RNA length distributions in nuclear fractionation and APEX-seq are very similar (Figure S3E).

General Features of the Human Transcriptome Revealed by APEX-Seq RNA Atlas

Our APEX-seq atlas reveals interesting patterns and features for the human transcriptome (Figure 3A). For >3,200 RNAs, we obtained high enrichment scores (log_2 fold change >0.75) in at least one of the nine locations. Unbiased clustering analysis revealed that RNAs broadly partition into four general localization categories (Figures 3A and 3D): (1) nuclear, (2) mitochondrial membrane and ER, (3) cytosol, and (4) the remaining (which includes ER lumen, mitochondrial matrix, and nuclear pore). Most transcripts further localized to just one or two locations within each category (Figures 3D and 3G; STAR Methods). Comparing mRNAs to long noncoding RNAs (lncRNAs) (Figures 3E and 3F), our dataset showed that the former mostly localize to one of the cytosolic or nuclear locations, while lncRNAs are predominantly nuclear, consistent with previous studies (Cabili et al., 2015).

We observed substantial overlap between OMM and ERM-associated transcriptomes (Figures 3H and 3I). Using more stringent cutoffs based on ROC analysis, we confirmed that two-thirds of RNAs are shared by OMM and ERM, with almost 95% of shared mRNAs encoding secreted proteins (Figures S3C and S3D). It may be that specific subsets of mRNAs are translated at mitochondria-ER contact sites (Friedman et al., 2011; Giacomello and Pellegrini, 2016; Valm et al., 2017).

We used our APEX-seq atlas to explore the relationship between protein localization and localization of its encoding mRNA, making use of existing data on protein subcellular localization (Thul et al., 2017). Our analysis (Figure 3J) reveals remarkable concordance between RNA and protein localization at steady state. For example, the ERM-proximal transcriptome preferentially codes for proteins that localize to the ER, Golgi, and vesicles, rather than proteins that localize to the nucleus, nucleolus, or cytosol. Less expectedly, our data also show that mRNAs enriched in nuclear locations tend to code for proteins enriched in nuclear speckles and nucleoplasm, but not the plasma membrane (Figures 3J, S3A, and S3B). This result is surprising if protein translation occurs exclusively in the cytosol. Alternatively, it has been suggested that mRNAs in the nucleus might serve as "reserve pools" that help to dampen gene-expression noise (Bahar Halpern et al., 2015; Battich et al., 2015; Hansen et al., 2018). We speculate that nuclear-destined proteins (Thul et al., 2017), which are highly enriched for nucleic-acid binding proteins (FDR < 5×10^{-13}, GO biological process) whose concentrations may have to be precisely tuned, may have mRNAs that are retained in nuclear subcompartments in order to better shield the amount of mRNA available for translation from noise.

The ability of our atlas to position endogenous RNAs with respect to distinct subcellular landmarks provides an exciting opportunity to test novel hypotheses concerning the relationship between RNA localization and function. For example, the atlas shows that *XIST* (X-inactive specific transcript), a nuclear lncRNA, is enriched at the nuclear lamina but not the nearby nuclear pore (Figure 3B). These findings are consistent with the known role of *XIST* in coating the inactive X chromosome in female cells (Penny et al., 1996), leading to transcriptional silencing and localization of the inactive X to the nuclear lamina (Chen et al., 2016). Another example is *IARS2* (mitochondrial isoleucyl tRNA synthetase 2, encoded by the nuclear genome), whose mRNA was identified by APEX-seq at the OMM (Figure 3C). Because *IARS2*'s protein product is known to reside in the mitochondrial matrix, the APEX-seq data suggest local translation of the mRNA at the OMM, a point we further explore in Figures 6 and 7.

Two other RNAs of note are *TUG1* and *NORAD*, lncRNAs localized by APEX-seq to both the ERM (validated by smFISH

(H) Circos plot showing the co-localization of RNAs to multiple locations.
(I) Transcripts overlapping in multiple locations.
(J) Heatmap showing the protein localization of the transcripts enriched by APEX-seq.
See also Figure S3.

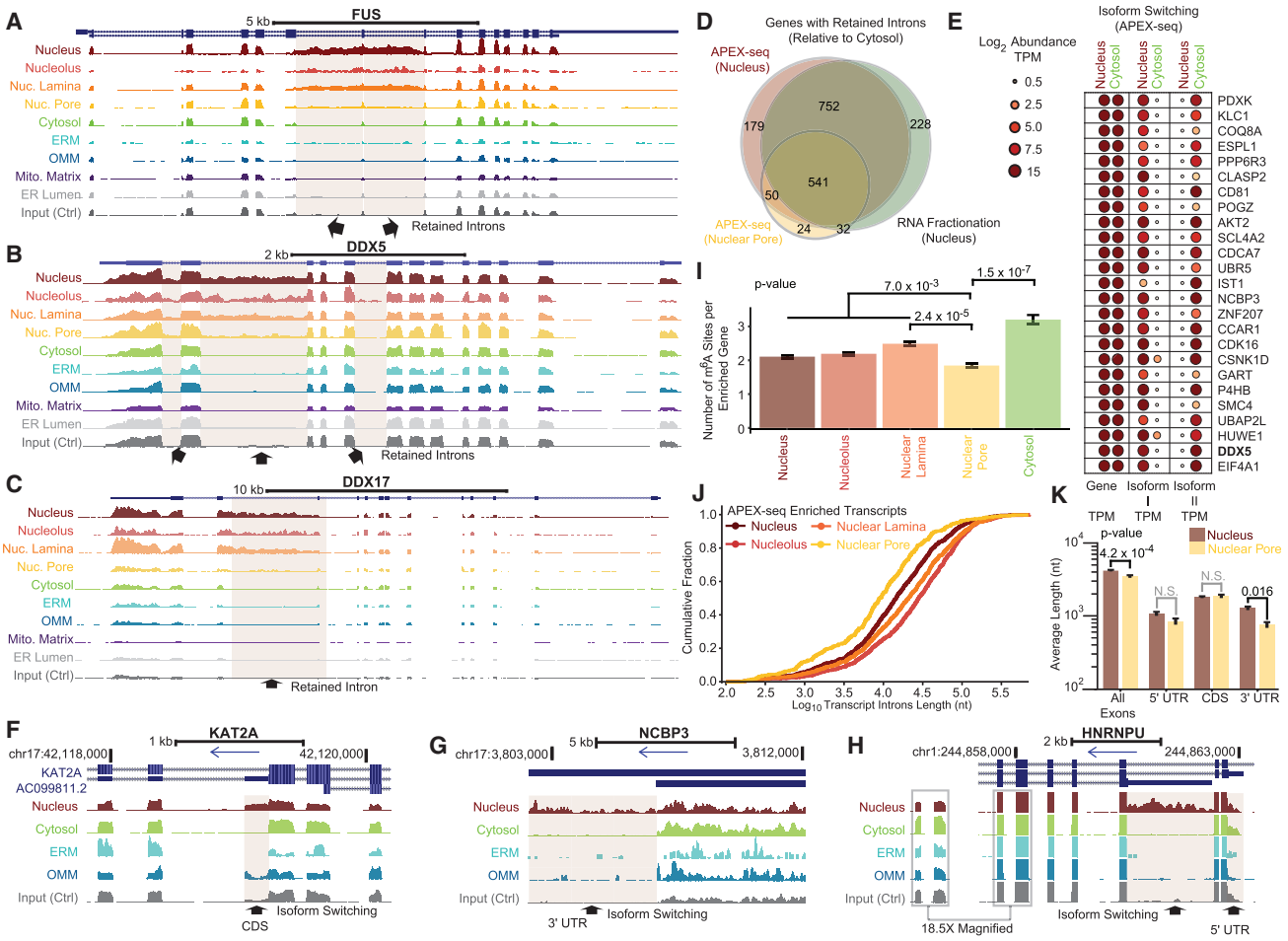


Figure 4. APEX-Seq Reveals Principles Related to RNA Isoforms and Introns

(A–C) The genome tracks of (A) *FUS* mRNA. (B) and (C) show the genome tracks of two other transcripts, *DDX5* and *DDX17*, with retained introns.

(D) Fractionation-seq (green) and nucleus APEX-seq (red) identify roughly the same genes with retained introns. The nuclear-pore APEX-seq transcriptome has fewer retained introns relative to the nucleus.

(E) Using APEX-seq, we identify transcripts that are highly abundant in both cytosol and nucleus at the gene level but switch isoforms at the transcript level. TPM, transcript per million.

(F–H) Browser tracks showing examples of isoform switching across nuclear and cytosolic locations for (F) *KAT2A* (lysine histone acetyltransferase 2A) in a putative coding sequence (CDS), (G) *NCBP3* (nuclear cap-binding protein subunit 3) in the 3′ UTR, and (H) *HNRNPU* (heterogenous nuclear ribonucleoprotein U) in the 5′ UTR, respectively. Arrows indicate direction of transcription.

(I) Number of m^6A present per transcript enriched by APEX-seq. High-confidence m^6A sites were obtained from the literature (Meyer et al., 2012). p values are from a Fisher's exact test.

(J) Cumulative distribution of the introns length for genes enriched by APEX-seq in the nuclear locations.

(K) Bar plots of average length of nuclear pore and nucleus enriched transcripts by mature transcript length, 5′ UTR, CDS (coding sequence) and 3′ UTR. p values are from a one-sided Mann-Whitney U test. Errors are SEM.

See also Figure S4.

imaging in Figures 2K and 2L) and the nucleus. While the majority (97%) of ERM APEX-seq-enriched species are mRNAs, our dataset highlights 31 noncoding RNAs, which are impossible to detect by ribosome profiling or ER fractionation-seq, because they are not translated.

APEX-Seq Reveals Differential Localization for Transcript Isoforms

Because APEX-seq is a sequencing-based methodology providing not only gene identity but full sequence details for each enriched RNA, we use it to support the hypothesis that different transcript isoforms of the same gene may localize to different regions of the cell (Mayr, 2017). For example, *FUS* (fused in sarcoma) mRNA, encoding a nuclear protein implicated in amyotrophic lateral sclerosis (ALS) and phase separation (Patel et al., 2015), shows intron retention within the nuclear locations, but not cytosolic ones (Figure 4A). Dead-BOX helicases 5 (*DDX5)* and 17 (*DDX17)* are additional examples of RNAs with retained introns (Figures 4B and 4C). The nuclear enrichment of retained introns was also observed in our fractionation-seq data

($r = 0.78$) although without the sub-nuclear resolution that APEX-seq provides. Interestingly, we observed that APEX at the nuclear pore enriched fewer transcripts with retained introns than APEX at other nuclear locations (Figures 4D and S4A–S4C), consistent with the role of the pore as a "gene gate" for RNA quality control.

In addition to retained introns, APEX-seq revealed a group of RNAs that show no gene-level subcellular localization differences but exhibit substantial spatial heterogeneity at the transcript-isoform level ("isoform switching"; Figures 4E and S4A–S4E). Two such examples are the mRNAs for the oncogene *AKT2* and the circadian rhythm gene *CSNK1D*, which show isoform switching between the nucleus and cytosol. In some cases, isoform switching extends to the 5′ UTR, 3′ UTR, and coding regions of transcripts (Figures 4F–4H). Overall, we find hundreds of genes with alternative 5′ and 3′ splice sites (Figures S4F and S4G). These results naturally nominate specific exons associated with each isoform for localization to specific subcellular locations, which in turn could affect downstream functions (Berkovits and Mayr, 2015).

Nuclear Pore as a Staging Area for RNA Export

RNA transcripts must pass through the nuclear pore to go from their production sites in the nucleus into the cytoplasm. Previous studies have suggested that the nuclear pore may act as a staging area for cytoplasm-destined transcripts (Wickramasinghe and Laskey, 2015). Our APEX-seq data reveal a striking similarity between RNAs enriched at the nuclear face of the nuclear pore (where APEX is expressed as a fusion to the pore-basket-binder SENP2 (Sentrin-specific protease 2) (Walther et al., 2001) and RNAs in the cytoplasm (Figure 3D), in contrast to RNAs from other nuclear locations (Figure 3A).

Our results support the prevailing view that the nucleoplasmic milieu (Blobel, 1985; Brown and Silver, 2007; Kim et al., 2018) of the pore has a critical role in mRNA surveillance, allowing only properly spliced and sorted transcripts ready for export to cytoplasm to congregate (while retaining partially spliced transcripts in the nucleus) (Figures 4A–4C).

m^6A Modification and RNA Length in Nuclear Pore Localization

While RNA processing for nuclear export is complex and highly regulated, the rate-limiting step for mRNA transport is believed to be access to and release from the nuclear pore complex (NPC) (Grünwald and Singer, 2010; Ma et al., 2013). *N6*-methyladenosine (m^6A) modification of pre-messenger RNAs has been reported as a "fast track" signal for nuclear export (Roundtree et al., 2017), while RNA length has been hypothesized as a feature influencing RNA export, with long RNAs taking more time to remodel and exit.

When we intersected nuclear-pore APEX-seq data with m^6A modification sites (Meyer et al., 2012), we found a significant depletion of m^6A in transcripts enriched near the pore, compared to nuclear lamina or the cytosol (Figure 4I). Our data support the hypothesis that m^6A-modified transcripts transit quickly through the NPC, leading to low biotinylation by APEX-seq. However, although transcripts at the pore had less m^6A than other nuclear locations, the transcript density of m^6A was not significantly different across these locations. Nonetheless, transcripts at both the pore and other nuclear locations had lower m^6A density (i.e., sites per kilobase) than the cytosol.

We also examined RNA length in our nuclear-pore APEX-seq data. We found that transcripts enriched at the pore tend to be shorter than transcripts at other nuclear locations. This inverse relationship between RNA length and nuclear pore APEX enrichment is significant both in the mature transcript and the introns only (Figures 4J, 4K, S4H, and S4K). For protein-coding transcripts, the 3′-UTR length is most predictive of nuclear pore APEX-seq enrichment (Figure 4K). A possible interpretation of our data is that longer RNAs pass more quickly through the pore, leading to lower APEX-seq enrichment, which could be the case if shorter RNAs assemble with fewer RNA-binding proteins (RBPs), including those necessary for recognition and passage through the pore.

Although different processes exist to export intronless mRNAs (Delaleau and Borden, 2015), we did not observe a significant difference in the proportion of intronless transcripts at the pore relative to other locations (Figure S4I).

RNA Repeats and Genomic Position Influence Sub-Nuclear RNA Localization

Repeat sequences make up a majority of the human genome (de Koning et al., 2011), with interspersed nuclear elements SINE (short) and LINE (long) containing retrotransposable (transposable via RNA intermediates) elements that can be deleterious when active and randomly moving to new genomic sites (Ichiyanagi, 2013). We observed enrichment of SINEs and LINEs within the nuclear locations (Figures 5A and S5A–S5D), with the highest enrichment of these elements in the nuclear lamina. The cytosolic locations and the nuclear pore showed no enrichment (Figure S5E). Given the known accumulation of transcription-repression machinery at the lamina (van Steensel and Belmont, 2017), our observations may help to explain the recent findings that LINE (L1) elements are epigenetically silenced (Padeken et al., 2015). Likewise, transcripts enriched at the nuclear lamina had lower expression level than other nuclear locations, consistent with the idea of heterochromatin deposition and gene silencing at lamina-associated domains (LADs) (Figures 5C and S5G–S5I).

Second, location of the DNA locus from which an RNA originates is believed to strongly dictate nuclear RNA location (Dekker et al., 2017), which we find support for. For example, previous work has shown that the nucleolus is enriched for DNA coding for rRNAs (van Koningsbruggen et al., 2010), while our APEX-seq atlas shows that rRNA repeat motifs (Wheeler et al., 2013) are highly enriched in the nucleolus but far less in the nuclear lamina or cytosol (Figure 5B). We also find that mRNA of genes residing in DNA nucleolus-associated domains (NADs) (Dillinger et al., 2017; van Koningsbruggen et al., 2010) are highly enriched in the nucleolus (odds ratio = 4.4; 95% confidence interval [CI] = 1.7–14) (Figures 5D and S5J). For DNA loci in LADs (Guelen et al., 2008), their corresponding RNA were enriched in the lamina APEX-seq (odds ratio = 11; 95% CI = 3.8–43) (Figures S5J–S5M).

Distinct Mechanisms of mRNA Localization to the OMM

Human mitochondrion contains >1,100 protein species (Calvo et al., 2016), only 13 of which are encoded by the mitochondrial

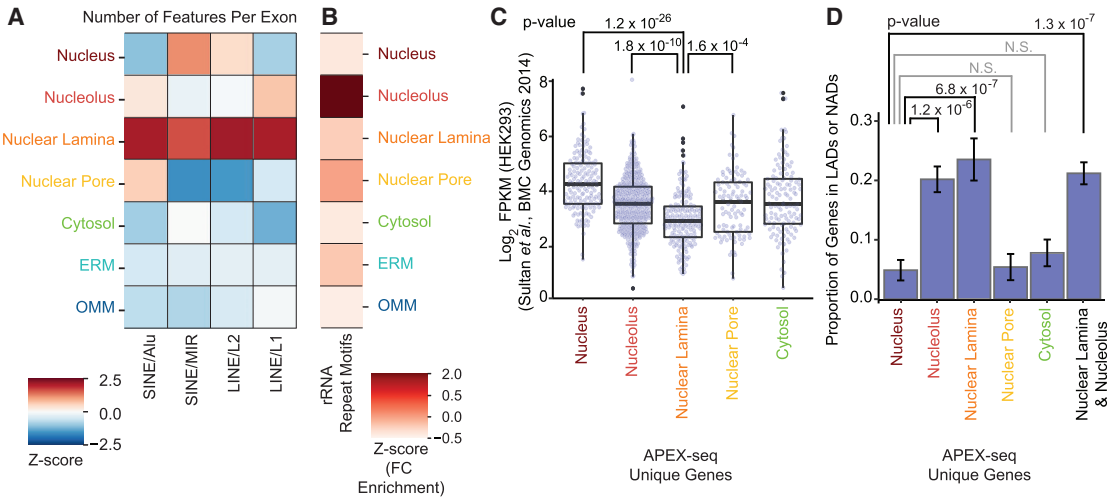


Figure 5. The Underlying Features of Nuclear RNA Localization

(A) Examination of retrotransposable elements in transcripts uniquely localizing to different locations show an enrichment of these elements in the nuclear-lamina transcriptome.

(B) Heatmap of *Z* score showing that transcripts localizing to the nucleolus are enriched in rRNA repeat motifs, relative to the nucleus.

(C) Within the nuclear locations, the nuclear-lamina-enriched transcripts have a lower abundance relative to both the nucleus and the nucleolus. p value is from a Mann-Whitney U test.

(D) Examination of the genes found in DNA lamina-associated domains (LADs) and nucleolus-associated domains (NADs) confirms that the corresponding transcriptomes are enriched for those genes. Here we restrict analysis to transcripts uniquely enriched in the respective locations. p values are from Fisher's exact tests.

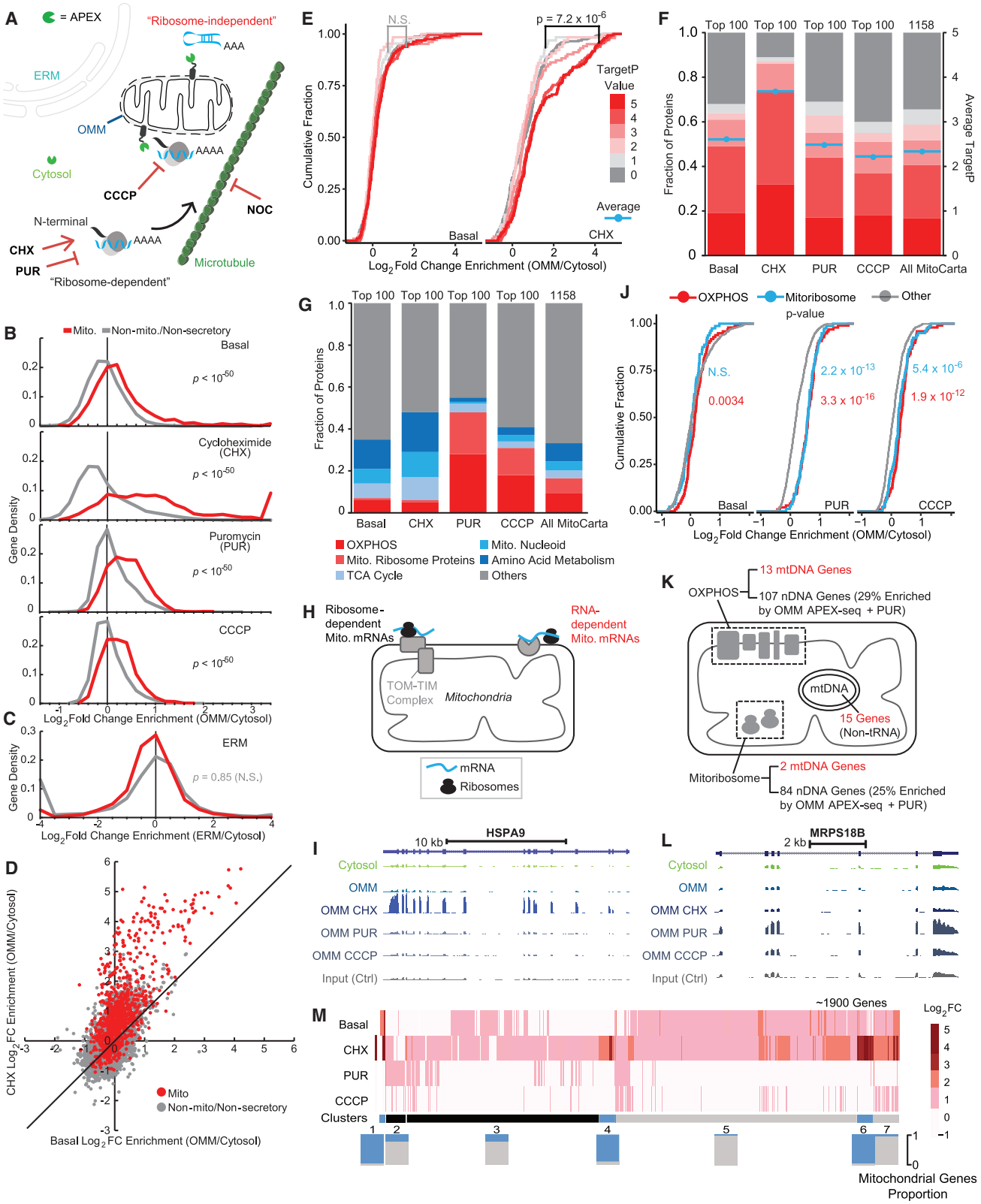
See also Figure S5.

genome (mtDNA) and translated within the organelle. The remainder are encoded by the nuclear genome and must be delivered to the mitochondrion after translation in the cytosol (Mercer et al., 2011). The identification of ribosomes at the OMM (Kellems et al., 1974, 1975) led to the hypothesis that some mRNAs encoding mitochondrial proteins may be locally translated at the OMM and co-translationally or post-translationally imported into the mitochondrion (Gold et al., 2017) (Figure 6A). However, at present little is known about the landscape of RNAs at the mammalian mitochondrial membrane, despite its importance for understanding mitochondrial biogenesis.

We mined our APEX-seq atlas for insights about mitochondria-proximal RNAs in human cells and found that the OMM compartment was enriched in mRNAs encoding mitochondrial proteins (Figure 3J). When plotted by OMM APEX-seq enrichment, we observed a significant increase in enrichment of nuclear-encoded mitochondrial genes over non-mitochondrial, non-secretory genes (Figure 6B; Table S4). By contrast, no increase in enrichment of mitochondrial genes was observed when RNAs were plotted by ERM APEX-seq enrichment score (Figure 6C). These results support the notion that mitochondrial transcripts accumulate at the OMM, possibly for the purpose of local protein translation. Examination of our OMM-enriched mRNAs did not reveal any pattern in terms of protein functional class or sub-mitochondrial localization of the encoded proteins. In an effort to further tease apart possible mRNA subpopulations that may be targeted to the OMM by different mechanisms, we repeated APEX-seq labeling under different perturbation conditions.

Taking advantage of the rapidity of APEX-seq tagging, we treated cells expressing OMM-APEX2 with cycloheximide (CHX), puromycin (PUR), or carbonyl cyanide m-chlorophenyl hydrazone (CCCP), prior to labeling (Figure S6A). CHX and PUR are both protein translation inhibitors but they work by different mechanisms. CHX stalls translation but preserves the mRNA-ribosome-nascent protein chain complex, while PUR dissociates mRNAs from ribosomes. CCCP abolishes the mitochondrial membrane potential and thereby stops membrane potential-dependent processes including TOM (translocase of outer membrane)/TIM-mediated import of mitochondrial proteins (Chacinska et al., 2009).

After treatment of cells with CHX, we observed a dramatic increase in the number of mitochondrial genes and their extent of OMM enrichment (Figures 6B and 6D), consistent with a model in which mRNA localization to the OMM can be regulated by the encoded protein's mitochondria-targeting sequence. As it emerges from the ribosome, the nascent peptide is localized to the OMM together with the still translating mRNA. Indeed, we found that the most-CHX-enriched mitochondrial genes have higher TargetP scores on average (Figures 6E–6G); TargetP is a measure of mitochondrial targeting potential (Emanuelsson et al., 2007). Hence, OMM APEX-seq following CHX appears to highlight a subpopulation of that may localize to the OMM in a ribosome-dependent fashion (Figure 6H). Figures 6I and S6C show the genome tracks of example mRNAs, *HSPA9* (mitochondria heat shock protein A9) and *MUT* (methylmalonyl-coA mutase), respectively, that display increased OMM localization upon CHX treatment.



(legend on next page)

Treatment of cells with PUR produced a pattern of enrichment distinct from CHX treatment. The vast majority of CHX-enriched mRNAs were no longer observed at the OMM, consistent with the hypothesis that the localization of these transcripts depends on an intact ribosome complex (Figures 6B and S6E). Nonetheless, a subpopulation of mRNAs remained clearly associated with the OMM after PUR; the top OMM-localized genes were not higher in TargetP, in contrast to CHX-enriched genes (Figure 6F). Functional class analysis revealed that PUR-enriched genes have a higher likelihood of encoding mitochondrial ribosome and oxidative phosphorylation (OXPHOS) components (Figures 6G, 6J, 6K, and S6F), which are the two complexes that require the coordinated assembly from the nuclear and mitochondrial genomes (Couvillion et al., 2016). Figures 6L and S6D show genome tracks of a representative mitochondrial ribosomal-protein gene, *MRPS18B* (28S ribosomal protein S18b), and OXPHOS gene, *NDUFB9* (NADH:ubiquinone oxidoreductase subunit B9), respectively. The PUR data thus suggest that a subpopulation of mRNAs associates with the OMM in a ribosome- and nascent-chain-independent fashion, perhaps by binding directly to a OMM-localized RNA binding protein (Figure 6H).

Upon treatment with the mitochondrial uncoupler CCCP, the genes enriched at the OMM are similar to PUR-enriched genes (Figures 6F, 6G, and 6J). CCCP-enriched genes must not depend on the mitochondrial membrane potential or mitochondrial protein import for their OMM localization. Perhaps by causing a reduction in interactions between the ribosome-mRNA-nascent chain complexes and TOM/TIM at the OMM, the association of ribosome-*independent* mRNAs with the OMM under CCCP becomes more readily apparent.

The availability of basal along with three ''drug perturbation'' OMM APEX-seq datasets enabled us to perform higher-order clustering analysis. Figure 6M shows transcripts that were enriched at the OMM in at least one condition. We find that RNAs cluster into groups based on their enrichment in CHX versus PUR, with some clusters strongly predictive of genes coding for mitochondrial proteins (Figure S6H). In particular, in clusters 1, 4, and 6 that included transcripts strongly enriched upon CHX treatment and depleted upon PUR treatment, >90% of RNAs (n = 128/140) code for mitochondrial proteins. 7 of the remaining 12 transcripts were pseudogenes, with at least 3 of the 5 mRNAs likely to be mitochondrial (Figures S6I–S6J) based on other studies (Mou et al., 2009; Pandey et al., 2017; Thul et al., 2017). Thus, OMM APEX-seq data could be used to predict whether certain genes will code for mitochondrial proteins.

Analysis of Motifs that Predict RNA Localization to the Mitochondrion

By using PUR and CHX treatments, we disentangled RNA populations that localize to the OMM via ribosome-dependent versus ribosome-independent mechanisms (Figure 6H). We next investigated two hypotheses: (1) that PUR-enriched mRNAs (''ribosome-independent'') possess specific RNA sequences that predict their OMM localization, and (2) that CHX-enriched mRNAs (''ribosome-dependent'') possess specific amino-acid features that predict OMM localization. To test these hypotheses, we first classified OMM-enriched transcripts as either ribosome-dependent or RNA dependent (if they localized to OMM under PUR) (Figure 7A). We trained a random-forest classification algorithm to predict localization of these two categories of transcripts to the OMM versus the ERM (which we used as ''background''), using 6-mers as RNA features (STAR Methods). The resulting classifier was much better at predicting localization of RNA-dependent transcripts relative to ribosome-dependent ones (Figure 7B). The converse result was obtained when using the corresponding N-terminal 100 amino acid peptide for training (Figures 7C and 7D), suggesting that the peptide sequence is more predictive for ribosome-dependent transcripts.

We looked further into the RNA features that may be predictive of OMM localization (Table S5) and found that the 5′ UTR was least important and the 3′ UTR most informative (Figures S7A and S7B). The most-important 6-mer sequences were G/U rich, with one of the other top hits being the poly(A)-signal

Figure 6. Distinct Subpopulations of mRNAs at the OMM

(A) Schematic diagram showing the mitochondria with all perturbations used in this study, including those that affect ribosomes (puromycin [PUR] and cycloheximide [CHX]), mitochondrial membrane potential (carbonyl cyanide m-chlorophenyl hydrazone [CCCP]), and microtubules (nocodazole [NOC]). RNA is shown in blue, ribosomes in gray, and microtubules in green.

(B) Gene density distribution of OMM APEX-seq enrichment under different conditions. p values are from Mann-Whitney U tests.

(C) Gene density distribution of ERM APEX-seq enrichment. Genes are categorized as in (B). p value is from a Mann-Whitney U test.

(D) Scatterplot of OMM APEX-seq $\log_2$ fold change comparing the basal and CHX conditions.

(E) Cumulative fraction of genes in different conditions by TargetP values. CHX treatment shows increased OMM targeting of genes with high TargetP values. Genes are categorized by their TargetP values (see STAR Methods) on a scale from 5 (strongest N-terminal mitochondrial targeting peptide) to 0 (no N-terminal mitochondrial targeting peptide). p values are from Kolmogorov-Smirnov (KS) test.

(F) Comparing the proportion of transcripts with different TargetP values and average TargetP value among top 100 mitochondrial genes enriched by OMM APEX-seq in cells under different conditions and all MitoCarta genes.

(G) Comparing the proportion of transcripts in different functional classes among top 100 mitochondrial genes enriched by OMM APEX-seq in cells under different conditions and all MitoCarta genes. Genes are functionally classified according to Gene Ontology.

(H) Model summarizing two distinct subpopulations of mitochondrial RNAs proximal to mitochondria.

(I) Browser tracks of a mitochondrial gene (*HSPA9*, targetP = 5) show increased enrichment by OMM-APEX upon CHX treatment.

(J) Cumulative fraction of OXPHOS and mitoribosome-related genes in different conditions. p values are from KS test.

(K) Scheme illustrating the coordinated assembly of respiratory chain complexes and mitoribosomes between the nuclear and mitochondrial genomes.

(L) Browser tracks of a mitochondrial ribosomal gene (*MRPS18B*) that show increased enrichment by OMM-APEX upon PUR or CCCP treatment.

(M) Heatmap of fold changes for transcripts enriched by OMM APEX-seq. Upon clustering based on the basal, CHX, and PUR conditions, we obtain clusters that are either strongly enriched or depleted in the corresponding mitochondrial proteins.

See also Figure S6 and Table S4.

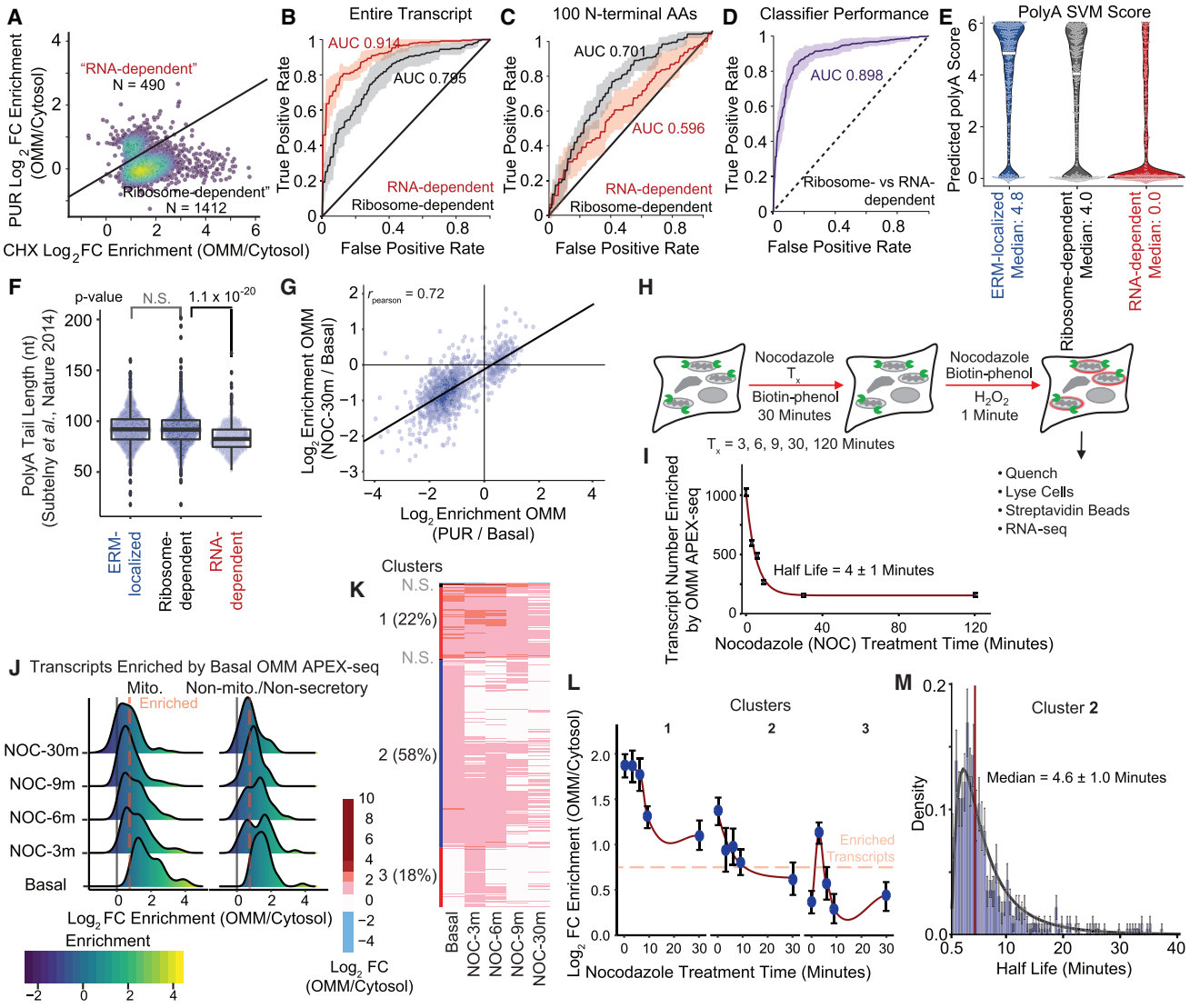


Figure 7. Features of Ribosome-Dependent and RNA-Dependent Transcripts at OMM

(A) Based on the effect of PUR and CHX, we binned genes from heatmap (Figure 6M) into two categories: ribosome dependent and RNA dependent.

(B) ROC curves from an unsupervised random-forest classifier that predicts transcript localization to OMM (versus ERM). To train the classifier, the transcript sequences were divided into 4,096 (= 4^6) 6-mers. Plotted is the mean performance (dark line) and the range from 10-fold cross-validation.

(C) Same as (B) but using the first 100 coding amino acids (aa) for training. Due to the much larger possible space of aa-variation, we used 3-mers (= 22^3 k-mers) instead of 6-mers for training.

(D) Similar model using 6-mer RNA sequences was used to classify transcripts as ribosome-dependent or RNA-dependent.

(E) Using the poly(A) SVM package, which predicts polyadenylation site scores, we find the RNA-dependent transcripts have low polyadenylation scores.

(F) Using a poly(A) tail-length dataset (Subtelny et al., 2014), we found RNA-dependent transcripts have shorter poly(A)-tail length relative to ribosome-dependent transcripts. p values are from Mann-Whitney U test.

(G) Correlation of fold change upon 30-min NOC treatment (where effect saturates) and the corresponding change upon PUR treatment. Changes are measured relative to basal conditions.

(H) Schematic diagram of the time-course APEX-seq protocol.

(I) Number of transcripts enriched by OMM-APEX-seq.

(J) Progressive depletion of basal OMM transcripts upon NOC treatment.

(K) Heatmap of genes enriched by APEX-seq in any of the time points. We clustered on the first 4 times points.

(L) Enrichment change as function of NOC treatment time for the three major clusters. Data are median fold change ±1 sigma.

(M) Half-lives for transcripts in Cluster 2.

See also Figure S7 and Tables S5 and S6.

sequence AAUAAA (Figure S7C). In support of our findings, the predicted poly(A) SVM score (a measure of poly(A)-site prediction) (Cheng et al., 2006) of RNA-dependent transcripts is substantially different from that of ribosome-dependent transcripts (Figure 7E). We also found that RNA-dependent OMM transcripts have significantly shorter poly(A)-tail lengths than ribosome-dependent transcripts, as well as shorter 3′ UTRs (Figures 7F and S7D). Altogether, our findings support the two hypotheses above and reveal specific RNA and protein features that are predictive of OMM localization.

Kinetics of RNA Transport to the Mitochondrion

Previous studies have suggested that RNA may arrive at the OMM via active microtubule-based transport (Buxbaum et al., 2015). To investigate this hypothesis, we repeated the OMM APEX-seq labeling after treating cells for various lengths of time with the microtubule-polymerization inhibitor nocodazole (NOC), which is known to inhibit transport (Reck-Peterson et al., 2018; Shen et al., 2018). We confirmed by imaging that NOC treatment does not perturb the localization of the OMM-APEX2 construct (Figure S7E). Figures 7H and 7I shows that 30 min of NOC led to a depletion of mRNAs at the OMM. The RNAs remaining at the OMM were more similar to those observed under PUR (r = 0.72) compared to those under CHX (r = 0.32) (Figures 7G and S7H). The selective disappearance of ribosome-dependent mRNAs from the OMM suggests that these mRNAs may utilize the cytoskeletal network to reach the OMM (Figure 6A).

Analysis of NOC time-course data (Figures 7H, 7I, S7F, and S7G; Table S6) showed that the majority of RNAs disappear rapidly from the OMM following NOC treatment. This decrease is observed for both mRNAs that encode mitochondrial proteins and other RNAs (Figure 7J). Further analysis resolved at least three patterns of responses to NOC (Figures 7K and 7L). The largest cluster shows rapid loss from the OMM with half-life dissociation data that could be fit by a log-normal distribution (Figure 7M), suggesting that many rate-limiting events could be involved. While further studies are needed to characterize these responses (as perturbing the cytoskeleton can have wide-ranging effects), our observations do showcase the power of rapid APEX-seq labeling to resolve dynamic transcriptome-wide RNA localization events.

DISCUSSION

With quantitative enrichment scores and detailed transcript profiles for over 25,000 distinct human RNA species across nine subcellular compartments, our study reveals patterns of RNA localization that give rise to a variety of biological hypotheses. APEX-seq yields RNA sequence information down to single-nucleotide resolution, thereby filling a critical gap in the landscape of RNA technologies. Our APEX-seq-derived atlas of transcriptome localization provides a comprehensive and precise delineation of RNA spatial organization in the living cell.

APEX-seq adds to arsenal of RNA localization methods while offering unique advantages. The first strength of APEX-seq is that labeling is performed in living cells, while membranes and macromolecular complexes are still intact. Second, APEX-seq can be used to analyze ''unpurifiable'' structures such as the nuclear lamina and OMM that are impossible to access via fractionation-based approaches. The third strength of APEX-seq is that it provides full sequence information for diverse classes of RNA transcripts, allowing transcript isoforms with distinct localization to be distinguished (Figures 4F–4H). Fourth, while ribosome profiling captures actively translating mRNA on polysomes, APEX-seq additionally detects lncRNAs, antisense RNAs (Figures 3E and 3F) and untranslated mRNAs not bound to ribosomes. Finally, the high spatiotemporal resolution sets APEX-seq apart from APEX-RIP, which loses spatial specificity in non-membrane enclosed regions (Figure 1D).

A disadvantage of APEX-seq is that it requires an APEX fusion construct to be recombinantly expressed in the cell of interest, which limits applicability to human tissue. Also, APEX-seq does not provide single-cell information like imaging-based methods. Finally, because labeling is performed in live cells, APEX-seq coverage will be fundamentally limited by the steric accessibility of RNAs in their native environment; RNAs that are buried within macromolecular complexes may not be tagged. These limitations suggest directions for future improvement.

We expect that APEX-seq will be broadly applicable to many organisms and cell types, just as APEX proteomics has been extended to flies (Chen et al., 2015a), worms (Reinke et al., 2017), yeast (Hwang and Espenshade, 2016), and neurons (Loh et al., 2016). APEX-seq could be fruitfully applied to polarized cells, neurons, or dynamic developmental systems. Future use of APEX-seq in conjunction with RNA-structure-mapping methods (Chin and Lécuyer, 2017; Spitale et al., 2015; Sun et al., 2019), RBP-occupancy atlases (Van Nostrand et al., 2016), and massively parallel reporter gene assays (Lubelsky and Ulitsky, 2018; Shukla et al., 2018) could shed light on the molecular basis of the spatial organization of RNA within cells.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mammalian cell culture
  - Generation of HEK293T cells stably expressing different APEX2 constructs
- METHOD DETAILS
  - APEX labeling in living cells
  - Immunofluorescence staining and fluorescence microscopy
  - RNA extraction for RT-qPCR or RNA-seq
  - APEX labeling streptavidin dot blot experiment
  - Horseradish peroxidase (HRP) *in vitro* labeling
  - Enrichment of biotinylated RNA
  - APEX-seq library preparation

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.05.027.

ACKNOWLEDGMENTS

We thank members of the Ting and Chang labs for analysis suggestions and critical reading of the manuscript, J. Coller and X. Ji for sequencing, and L. Mateo for assistance with microscopy. This work was supported by NIH (R01-CA186568-1 to A.Y.T.; R01-HG004361 and P50-HG007735 to H.Y.C.; S10OD018220 to Stanford Functional Genomics Facility), Burroughs Wellcome (CASI grant to A.N.B.), and the Chan Zuckerberg Biohub (to A.Y.T.). F.M.F. was supported by the NIH T32 SGTP and the Arnold. O. Beckman Postdoctoral Fellowships. S.H. was supported by the Stanford Bio-X Bowes Fellowship. H.Y.C. is an Investigator of HHMI. A.Y.T. is a Chan Zuckerberg Biohub Investigator.

AUTHOR CONTRIBUTIONS

F.M.F., P.K., H.Y.C., and A.Y.T. conceived the project. F.M.F., S.H., P.K., H.Y.C., and A.Y.T. designed experiments. F.M.F., S.H., and P.K. performed all experiments, unless otherwise noted. F.M.F. designed and carried out sequencing experiments. S.H., A.N.B., and A.Y.T. designed and carried out sequential FISH experiments. F.M.F., S.H., K.R.P., P.K., J.X., and A.Y.T. analyzed data. F.M.F., S.H., H.Y.C., and A.Y.T. wrote the paper with input from all authors. H.Y.C. and A.Y.T. jointly supervised work and acquired funding.

DECLARATION OF INTERESTS

A.Y.T., P.K., H.Y.C., and F.M.F. have filed a patent application covering aspects of this work (Patent Application Number US 2017/0226561). H.Y.C. is a co-founder and advisor of Accent Therapeutics. H.Y.C. is an advisor of 10X Genomics and Spring Discovery.

Received: August 31, 2018
Revised: December 31, 2018
Accepted: May 14, 2019
Published: June 20, 2019

REFERENCES

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. *22*, 2008–2017.

Anders, S., Pyl, P.T., and Huber, W. (2014). HTSeq - A Python Framework to Work with High-Throughput Sequencing Data (Cold Spring Harbor Laboratory).

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep. *13*, 2653–2662.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. Cell *163*, 1596–1610.

Benhalevy, D., Anastasakis, D.G., and Hafner, M. (2018). Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. Nat. Methods *15*, 1074–1082.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363–367.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. Mol. Cell *2*, 437–445.

Blobel, G. (1985). Gene gating: a hypothesis. Proc. Natl. Acad. Sci. USA *82*, 8527–8529.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Brown, C.R., and Silver, P.A. (2007). Transcriptional regulation at the nuclear pore complex. Curr. Opin. Genet. Dev. *17*, 100–106.

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.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. *16*, 20.

Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. *44* (D1), D1251–D1257.

Chacinska, A., Koehler, C.M., Milenkovic, D., Lithgow, T., and Pfanner, N. (2009). Importing mitochondrial proteins: machineries and mechanisms. Cell *138*, 628–644.

Chen, C.L., Hu, Y., Udeshi, N.D., Lau, T.Y., Wirtz-Peitz, F., He, L., Ting, A.Y., Carr, S.A., and Perrimon, N. (2015a). Proteomic mapping in live Drosophila tissues using an engineered ascorbate peroxidase. Proc. Natl. Acad. Sci. USA *112*, 12093–12098.

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

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.

Cheng, Y., Miura, R.M., and Tian, B. (2006). Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics *22*, 2320–2325.

Chin, A., and Lécuyer, E. (2017). RNA localization: Making its way to the center stage. Biochim. Biophys. Acta, Gen. Subj. *1861* (11 Pt B), 2956–2970.

Couvillion, M.T., Soto, I.C., Shipkovenska, G., and Churchman, L.S. (2016). Synchronized mitochondrial and cytosolic translation programs. Nature *533*, 499–503.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. *7*, e1002384.

Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O'Shea, C.C., Park, P.J., Ren, B., et al.; 4D Nucleome Network (2017). The 4D nucleome project. Nature *549*, 219–226.

Delaleau, M., and Borden, K.L. (2015). Multiple export mechanisms for mRNAs. Cells *4*, 452–473.

Dillinger, S., Straub, T., and Németh, A. (2017). Nucleolus association of chromosomal domains is largely maintained in cellular senescence despite massive nuclear reorganisation. PLoS ONE *12*, e0178821.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq reads with STAR. Curr. Protoc. Bioinformatics *51*, 1–19.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.

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

Fasken, M.B., and Corbett, A.H. (2009). Mechanisms of nuclear mRNA quality control. RNA Biol. *6*, 237–241.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A., Chang, H.Y., and Ting, A.Y. (2019). APEX-seq: RNA subcellular localization by proximity labeling. Protocol Exchange. Published online May 15, 2019. https://doi.org/10.21203/rs.2.1857/v1.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585–590.

Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., and Chang, H.Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc. *11*, 273–290.

Fox, C.H., Johnson, F.B., Whiting, J., and Roller, P.P. (1985). Formaldehyde fixation. J. Histochem. Cytochem. *33*, 845–853.

Friedman, J.R., Lackner, L.L., West, M., DiBenedetto, J.R., Nunnari, J., and Voeltz, G.K. (2011). ER tubules mark sites of mitochondrial division. Science *334*, 358–362.

Gagnon, K.T., Li, L., Janowski, B.A., and Corey, D.R. (2014). Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. Nat. Protoc. *9*, 2045–2060.

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.

Giacomello, M., and Pellegrini, L. (2016). The coming of age of the mitochondria-ER contact: a matter of thickness. Cell Death Differ. *23*, 1417–1427.

Gold, V.A., Chroscicki, P., Bragoszewski, P., and Chacinska, A. (2017). Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryotomography. EMBO Rep. *18*, 1786–1800.

Grünwald, D., and Singer, R.H. (2010). In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. Nature *467*, 604–607.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811–2812.

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

Han, S., Udeshi, N.D., Deerinck, T.J., Svinkina, T., Ellisman, M.H., Carr, S.A., and Ting, A.Y. (2017). Proximity biotinylation as a method for mapping proteins associated with mtDNA in living cells. Cell Chem. Biol. *24*, 404–414.

Han, S., Li, J., and Ting, A.Y. (2018). Proximity labeling: spatially resolved proteomic mapping for neurobiology. Curr. Opin. Neurobiol. *50*, 17–23.

Hansen, M.M.K., Desai, R.V., Simpson, M.L., and Weinberger, L.S. (2018). Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. Cell Systems *7*, 384–397.

Hensen, F., Moretton, A., van Esveld, S., Farge, G., and Spelbrink, J.N. (2018). The mitochondrial outer-membrane location of the EXD2 exonuclease contradicts its direct role in nuclear DNA repair. Sci. Rep. *8*, 5368.

Hung, V., Zou, P., Rhee, H.-W., Udeshi, N.D., Cracan, V., Svinkina, T., Carr, S.A., Mootha, V.K., and Ting, A.Y. (2014). Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. Mol. Cell *55*, 332–341.

Hung, V., Udeshi, N.D., Lam, S.S., Loh, K.H., Cox, K.J., Pedram, K., Carr, S.A., and Ting, A.Y. (2016). Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. Nat. Protoc. *11*, 456–475.

Hung, V., Lam, S.S., Udeshi, N.D., Svinkina, T., Guzman, G., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2017). Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. eLife *6*. Published online April 25, 2017. https://doi.org/10.7554/eLife.24463.

Hwang, J., and Espenshade, P.J. (2016). Proximity-dependent biotin labelling in yeast using the engineered ascorbate peroxidase APEX2. Biochem. J. *473*, 2463–2469.

Ichiyanagi, K. (2013). Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. Genes Genet. Syst. *88*, 19–29.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

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

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*. Published online December 14, 2017. https://doi.org/10.7554/eLife.29224.

Käll, 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.

Kellems, R.E., Allison, V.F., and Butow, R.A. (1974). Cytoplasmic type 80 S ribosomes associated with yeast mitochondria. II. Evidence for the association of cytoplasmic ribosomes with the outer mitochondrial membrane in situ. J. Biol. Chem. *249*, 3297–3303.

Kellems, R.E., Allison, V.F., and Butow, R.A. (1975). Cytoplasmic type 80S ribosomes associated with yeast mitochondria. IV. Attachment of ribosomes to the outer membrane of isolated mitochondria. J. Cell Biol. *65*, 1–14.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

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.

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., et al. (2018). Integrative structure and functional anatomy of a nuclear pore complex. Nature *555*, 475–482.

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.

Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K., and Ting, A.Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat. Methods *12*, 51–54.

Lee, B., Flynn, R.A., Kadina, A., Guo, J.K., Kool, E.T., and Chang, H.Y. (2017). Comparison of SHAPE reagents for mapping RNA structures inside living cells. RNA *23*, 169–174.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. IEEE Trans. Vis. Comput. Graph. *20*, 1983–1992.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Linden, A. (2006). Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J. Eval. Clin. Pract. *12*, 132–139.

Loh, K.H., Stawski, P.S., Draycott, A.S., Udeshi, N.D., Lehrman, E.K., Wilton, D.K., Svinkina, T., Deerinck, T.J., Ellisman, M.H., Stevens, B., et al. (2016). Proteomic analysis of unbounded cellular compartments: synaptic clefts. Cell *166*, 1295–1307.

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.

Lubelsky, Y., and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. Nature *555*, 107–111.

Ma, J., Liu, Z., Michelotti, N., Pitchiaya, S., Veerapaneni, R., Androsavich, J.R., Walter, N.G., and Yang, W. (2013). High-resolution three-dimensional mapping of mRNA export through the nuclear pore. Nat. Commun. *4*, 2414.

Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. (2011). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. *39*, D52–D57.

Mayr, C. (2017). Regulation by 3′-Untranslated Regions. Annu. Rev. Genet. *51*, 171–194.

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.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566.

Moffitt, J.R., and Zhuang, X. (2016). RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). Methods Enzymol. *572*, 1–49.

Mortensen, A., and Skibsted, L.H. (1997). Importance of carotenoid structure in radical-scavenging reactions. J. Agric. Food Chem. *45*, 2970–2977.

Mou, Z., Tapper, A.R., and Gardner, P.D. (2009). The armadillo repeat-containing protein, ARMCX3, physically and functionally interacts with the developmental regulatory factor Sox10. J. Biol. Chem. *284*, 13629–13640.

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Längst, G. (2010). Initial genomics of the human nucleolus. PLoS Genet. *6*, e1000889.

Padeken, J., Zeller, P., and Gasser, S.M. (2015). Repeat DNA in genome organization and stability. Curr. Opin. Genet. Dev. *31*, 12–19.

Pandey, R.R., Homolka, D., Chen, K.M., Sachidanandam, R., Fauvarque, M.O., and Pillai, R.S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. PLoS Genet. *13*, e1006956.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A liquid-to-solid phase transition of the ALS Protein FUS accelerated by disease mutation. Cell *162*, 1066–1077.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

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.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *35*, D61–D65.

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

Reck-Peterson, S.L., Redwine, W.B., Vale, R.D., and Carter, A.P. (2018). The cytoplasmic dynein transport machinery and its many cargoes. Nat. Rev. Mol. Cell Biol. *19*, 382–398.

Reid, D.W., and Nicchitta, C.V. (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J. Biol. Chem. *287*, 5518–5527.

Reid, D.W., and Nicchitta, C.V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *16*, 221–231.

Reinke, A.W., Mak, R., Troemel, E.R., and Bennett, E.J. (2017). In vivo mapping of tissue- and subcellular-specific proteomes in *Caenorhabditis elegans*. Sci. Adv. *3*, e1602426.

Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science *339*, 1328–1331.

Roundtree, I.A., Luo, G.Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., et al. (2017). YTHDC1 mediates nuclear export of N^6-methyladenosine methylated mRNAs. eLife *6*. Published online October 6, 2017. https://doi.org/10.7554/eLife.31311.

Sadowski, P.G., Groen, A.J., Dupree, P., and Lilley, K.S. (2008). Sub-cellular localization of membrane proteins. Proteomics *8*, 3991–4011.

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.

Schnell, U., Dijk, F., Sjollema, K.A., and Giepmans, B.N.G. (2012). Immunolabeling artifacts and the need for live-cell imaging. Nat. Methods *9*, 152–158.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron *92*, 342–357.

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. USA *111*, E5593–E5601.

Shen, J., Zhang, J.H., Xiao, H., Wu, J.M., He, K.M., Lv, Z.Z., Li, Z.J., Xu, M., and Zhang, Y.Y. (2018). Mitochondria are transported along microtubules in membrane nanotubes to rescue distressed cardiomyocytes from apoptosis. Cell Death Dis. *9*, 81.

Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry, R.A., Maass, P.G., and Rinn, J.L. (2018). High-throughput identification of RNA nuclear enrichment sequences. EMBO J. *37*. Published online March 15, 2018. https://doi.org/10.15252/embj.201798452.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., and Chang, H.Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H., and Yaspo, M.L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics *15*, 675.

Sun, L., Fazal, F.M., Li, P., Broughton, J.P., Lee, B., Tang, L., Huang, W., Kool, E.T., Chang, H.Y., and Zhang, Q.C. (2019). RNA structure maps across mammalian cellular compartments. Nat. Struct. Mol. Biol. *26*, 322–330.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. *13*, 2129–2141.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*. Published online May 26, 2017. https://doi.org/10.1126/science.aal3321.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. B *63*, 411–423.

Valm, A.M., Cohen, S., Legant, W.R., Melunis, J., Hershberg, U., Wait, E., Cohen, A.R., Davidson, M.W., Betzig, E., and Lippincott-Schwartz, J. (2017). Applying systems-level spectral imaging and analysis to reveal the organelle interactome. Nature *546*, 162–167.

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

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G.J., Ariyurek, Y., den Dunnen, J.T., and Lamond, A.I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol. Biol. Cell *21*, 3735–3748.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

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

Walther, T.C., Fornerod, M., Pickersgill, H., Goldberg, M., Allen, T.D., and Mattaj, I.W. (2001). The nucleoporin Nup153 is required for nuclear pore basket formation, nuclear pore complex anchoring and import of a subset of nuclear proteins. EMBO J. *20*, 5703–5714.

Weil, T.T., Parton, R.M., and Davis, I. (2010). Making the message clear: visualizing mRNA localization. Trends Cell Biol. *20*, 380–390.

Wheeler, T.J., Clements, J., Eddy, S.R., Hubley, R., Jones, T.A., Jurka, J., Smit, A.F., and Finn, R.D. (2013). Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic Acids Res. *41*, D70–D82.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer).

Wickramasinghe, V.O., and Laskey, R.A. (2015). Control of mammalian gene expression by selective mRNA export. Nat. Rev. Mol. Cell Biol. *16*, 431–442.

Williams, C.C., Jan, C.H., and Weissman, J.S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science *346*, 748–751.

Wishart, J.F., and Rao, B.S.M. (2010). Recent Trends in Radiation Chemistry (World Scientific).

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415.