



Published in final edited form as:

*Nat Methods*. 2018 December ; 15(12): 1074–1082. doi:10.1038/s41592-018-0220-y.

## Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments

Daniel Benhalevy, Dimitrios Anastasakis, and Markus Hafner\*

Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Disease, Bethesda, Maryland, USA

### Abstract

Methods to systematically study subcellular RNA localization are limited and lagging behind proteomic tools. Here, we combined APEX2-mediated proximity biotinylation of proteins with photoactivatable ribonucleoside-enhanced crosslinking to simultaneously profile the proteome, as well as the transcriptome bound by RNA-binding proteins in any given subcellular compartment. Our approach is fractionation-independent and enables to study the localization of RNA processing intermediates, as well as the identification of regulatory RNA *cis*-acting elements occupied by proteins in a cellular compartment-specific manner. We applied Proximity-CLIP to study RNA and protein in the nucleus, cytoplasm and at cell-cell interfaces. Among other insights, we observed frequent transcriptional readthrough continuing for several kilobases downstream of the canonical cleavage and polyadenylation site and a differential RBP occupancy pattern for mRNAs in the nucleus and cytoplasm. Surprisingly, mRNAs localized to cell-cell interfaces often encoded regulatory proteins and contained protein-occupied CUG sequence elements in their 3' untranslated region.

### Introduction

The distribution of messenger RNAs (mRNAs) and other RNA transcripts to specific subcellular locations has been widely studied using biochemical fractionation and/or hybridization and microscopy-based approaches. These studies revealed that a substantial portion of mRNAs and other transcripts are differentially localized<sup>2</sup> and that this localization is regulated and evolutionarily conserved<sup>5</sup> and likely plays a role in shaping gene expression<sup>6</sup>. The combination of next-generation sequencing and biochemical fractionation

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#termshttp://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#termshttp://www.nature.com/authors/editorial_policies/license.html#terms)

\*Correspondence to M.H. (markus.hafner@nih.gov).

#### AUTHOR CONTRIBUTIONS

Conceptualization, D.B. and M.H.; Methodology, D.B. and D.A.; Data Analysis, D.B. and D.A.; Writing - Original Draft, D.B. and M.H.; Writing - Review & Editing, D.B., D.A., M.H.; Funding Acquisition, M.H.; Resources, M.H.; Supervision, M.H.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

#### DATA AVAILABILITY

NGS sequence data has been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under accession number GSE110380. Raw blots used to produce Figure 2bc and Supplementary Figures 1a and 3a are provided as Suppl. Fig. 12. See the Life Sciences Reporting Summary file for more details.

allows for systems-level profiling of RNA localization, but was limited to the analysis of a small number of compartments including the nucleus, cytoplasm, mitochondria, and large structures such as neuronal dendrites or axons<sup>1</sup>.

Only recently, the more advanced tools available for high-throughput protein localization studies are being adapted for determination of RNA localization. Kaewsapsak et al. used APEX2-mediated proximity biotinylation, followed by formaldehyde crosslinking to catalogue mRNA and long non-coding RNAs (lncRNA) in the nucleus, mitochondria, and endoplasmic reticulum. The engineered APEX2 peroxidase can oxidize biotin-phenol in the presence of hydrogen peroxide to generate rapidly decaying biotin-phenoxy radicals ( $t_{1/2} < 1$  ms) that can biotinylate proximal proteins at aromatic amino acid side-chains. By fusion to localization signals, APEX2 is targeted to specific cellular compartments and biotinylates the localized proteome that can then be easily isolated by affinity chromatography and analyzed using mass spectrometry (MS). This approach has been successfully used to quantify the localized proteome of multiple subcellular compartments (reviewed in). Here, we present Proximity-CLIP, a method that combines well-established compartment-specific protein biotinylation with UV crosslinking, which irreversibly crosslinks RNA with RNA binding proteins (RBPs) in intact cells. Proximity-CLIP allows for (1.) the detection of the localized “RBPome”, (2.) quantification of the localized transcriptome, and (3.) the identification of RBP-occupied RNA loci including *cis*-acting regulatory elements and sites of RNA metabolism. In a proof-of-principle study, we applied Proximity-CLIP to three cellular compartments in HEK293 cells: the nucleus, the cytoplasm, and cell-cell interface.

## Results

### The Proximity-CLIP approach to identify localized ribonucleoproteins and RNAs

We reasoned that covalent crosslinking of RNA with interacting proteins by photoreactive ribonucleoside-enhanced crosslinking, combined with compartment specific protein biotinylation using the APEX2 system will allow to identify the transcripts localized in that compartment and we termed this approach Proximity-CLIP. Our approach relies on the well-supported assumptions that most cellular RNAs are protein-bound throughout their life cycle (Fig. 1a) and that RBPs from different subcellular compartments remain amenable to UV-crosslinking.

Proximity-CLIP comprises the following steps: (1.) 4-thiouridine (4SU) labeling of RNAs in living cells expressing specifically localized APEX2; (2.) biotinylation of APEX2-proximate proteins for 1 min and reaction quenching; (3.) *in-vivo* crosslinking of RNA and proteins using UV light ( $\lambda > 310$  nm) during the quenching step; and (4.) isolation of localized, biotinylated, and crosslinked ribonucleoprotein (RNP) complexes by affinity chromatography (Fig. 1b).

The covalent nature of the interactions between biotin, RBPs, and RNA renders the RNP complexes resistant to stringent purification steps, maximizing the signal-to-noise ratio in the downstream high-throughput proteomic and transcriptomic analyses (Fig. 1b). Proximity biotinylation eliminates the need for cell fractionation schemes and allows for the isolation of compartments inaccessible to biochemical purification. Nevertheless, it should be noted

that protein and RNA yield from Proximity-CLIP is expected to correlate with the input material abundance, reducing the signal-to-noise ratio for subcellular compartments containing small amounts of protein and/or RNA.

Proximity-CLIP allows for: (1.) the determination of the localized proteome by MS - as previously established by Ting et al. - which includes the RBPome that can be defined by comparison with published compendia of RBPs-; (2.) the profiling of localized transcripts using RNAseq; and (3.) the identification and quantification of RBP-occupied *cis*-acting elements on transcripts, by sequencing of cDNA libraries from RNase-resistant footprints (Fig. 1c). UV-crosslinking of 4SU-labeled RNA to interacting proteins results in a characteristic T-to-C mutation in the corresponding cDNA libraries. This feature allows for efficient computational removal of contaminating sequences derived from non-crosslinked fragments of abundant cellular RNAs, further increasing the specificity of Proximity-CLIP.

### Proximity-CLIP identifies nuclear and cytoplasmic RNAs and proteins

As proof-of-principle we first applied Proximity-CLIP to the cytoplasmic and nuclear compartments of human cells and generated stable HEK293 cells inducibly expressing V5-tagged APEX2 constructs fused either to a nuclear export signal (NES), or to histone H2B (Fig. 2a). As expected, immunofluorescence revealed protein biotinylation in the cytoplasm or the nucleus, respectively, upon expression of APEX2-NES or H2B-APEX2 only after treatment with both BP and H<sub>2</sub>O<sub>2</sub>, (Fig. 2a). Immunoblot analysis of lysates from 4SU-treated and UV-crosslinked cells also showed biotinylation of endogenous proteins in an APEX2-, BP- and H<sub>2</sub>O<sub>2</sub>-dependent manner (Fig. 2b,c). Biotinylated proteins, including RNPs were then isolated by streptavidin affinity chromatography (Suppl. Fig. 1a). We profiled the nuclear and cytoplasmic proteome by tryptic digestion of the bead-immobilized material and analysis of the eluting peptides by MS (Suppl. Data 1). To test whether Proximity-CLIP identified proteins from the correct compartment, we performed functional enrichment analysis of proteins biotinylation by H2B-APEX2 and APEX2-NES, respectively (Suppl. Data 1). As expected, proteins biotinylation by H2B-APEX2 were categorized as belonging to the “nucleus”, “nucleoplasm”, or to transcription-related processes. Stringent filtering and comparison relative to cytoplasmic proteins resulted a list of 137 highly enriched nuclear proteins, of which 86 were RBPs that belonged to the mRNA processing or export machinery (Fig. 2d), underscoring the centrality of RNA metabolism in nuclear processes. In contrast, while all enriched proteins biotinylation by APEX2-NES were known to be cytoplasmic, less annotation terms significantly enriched, likely reflecting their functional diversity (Fig. 2e). Finally, we verified that our proteome profiles correlated with a previously reported MS-based study of nuclear and cytoplasmic proteins collected using other methods and cell lines (Suppl. Fig. 1b-e). In summary, Proximity-CLIP reliably identified compartmentalized proteins, including RBPs possibly pulling down crosslinked to RNA.

We then aimed to probe whether the capture of nuclear and cytoplasmic RNAs by Proximity-CLIP would recapitulate results obtained by biochemical fractionation of cytoplasm and nucleus. We eluted RNA using Proteinase K from streptavidin-immobilized and crosslinked RNPs respectively biotinylation by APEX2-NES and H2B-APEX2, followed

by standard RNAseq library preparation and sequencing (Suppl. Data 2). Our RNAseq results of both cytoplasmic and nuclear Proximity-CLIP correlated well with the RNA profiles from corresponding cell fractions ( $r_s=0.72$  and  $r_s=0.80$ , respectively, Fig. 2f,g), demonstrating the potential of Proximity-CLIP to complement or substitute traditional biochemical fractionation.

Proximity-CLIP may provide global insights into protein occupancy on RNA. We ribonuclease treated crosslinked and streptavidin-immobilized nuclear and cytoplasmic RNPs, and sequenced isolated protein-protected RNA segments. We used the PARalyzer software to determine RBP binding sites or footprints that consisted of overlapping reads that contained T-to-C mutations diagnostic of the crosslinking event at higher frequencies than expected by chance (Suppl. Data 3). In total we identified 156,309 and 49,972 footprints in the nucleus and the cytoplasm, respectively (Fig. 2h). Footprint annotation categories accurately reflected their subcellular origin: of the 41,203 RBP footprints on mRNA in the cytoplasm, 94% were found on mature mRNA, while of the 125,265 mRNA footprints in the nucleus, 43% resided in introns, similar to typical interactome profiles of nuclear or cytoplasmic RBPs (Fig. 2h).

### **Proximity-CLIP reveals features of nuclear and cytoplasmic transcripts.**

Proximity-CLIP allowed us to test whether protein-coding transcripts interact with RBPs in a different manner prior to nuclear export and in the cytoplasm. Metagene analysis of the coverage of nuclear and cytoplasmic RBP-protected footprints showed a decay in coverage near the transcription start site (TSS) and more pronounced towards the 3' end of the mRNA 3' untranslated region (UTR) (Fig. 3a) that was likely due to alternative TSS usage and cleavage and polyadenylation (CPA) sites<sup>19,20</sup>, respectively. Consistent with a previous study analyzing protein occupancy on polyadenylated RNA, we found that cytoplasmic mRNAs are predominantly bound by RBPs at their 3'UTRs. Interestingly however, this was not the case for nuclear mRNAs; the highest density of nuclear RBP footprints was found in their 5' UTRs, rather than 3' UTR, in agreement with the dominant roles of the 5' cap binding complex in mRNA nuclear export regulation-. This pattern was not observed previously by studying RBP binding to total mRNA, probably due to the lower relative abundance of nuclear versus cytoplasmic mRNA.

### **Proximity-CLIP detects mRNA processing intermediates**

Next, we asked whether Proximity-CLIP captured nuclear RNA processing intermediates, such as intronic sequences, or immature pre-mRNA 3' ends. As expected, cytoplasmic mRNAs were depleted of intronic sequences, while nuclear RBP footprints exhibited even coverage of intronic sequences upstream and downstream of exons (Fig. 3b). We found footprint coverage upstream of the annotated splicing branchpoints in the nuclear Proximity-CLIP footprints, and a dramatic increase in footprints 20–50 nt downstream of the branchpoint, in good accordance with the known distances of the branchpoint to the downstream exon (Fig. 3c). We also observed a slight dip in nuclear RBP footprint coverage around the branchpoint itself, which may be due to RNA-RNA interactions with U2 snRNA, precluding interaction and crosslinking to RBPs.

We next tested whether we could observe mRNA 3' end formation that involves cotranscriptional endonucleolytic cleavage and subsequent polyadenylation at the CPA site. As expected, footprint coverage from the cytoplasmic Proximity-CLIP experiment was virtually undetectable after the annotated CPA sites, while, nuclear RNA footprints were detected downstream of the CPA site (Fig. 3d). These sequences may represent the residual uncapped transcript emerging from PolIII after CPA that needs to be degraded by XRN2 before PolIII can disassociate from chromatin according to the torpedo model of transcription termination. Consistently, a decay in coverage was apparent with distance from the CPA site, nevertheless, several kilobases downstream of the CPA we still detected RBP protected footprints (Suppl. Fig. 2a,b), in agreement with recently reported widespread transcriptional readthrough<sup>28,29</sup>.

Finally, we focused on the TSS to test whether Proximity-CLIP can recapitulate transcriptional features. By separating sense- and antisense footprints around the TSS, we were able to observe bidirectional divergent transcription products in nuclear footprints, with only the sense footprints found in the cytoplasm (Fig. 3e, Suppl. Fig. 2c), reflecting a proposed bidirectional nature of most promoters and a rapid degradation of the antisense transcript-. We also identified some coverage of sense-oriented nuclear footprints directly upstream of the TSS (Suppl. Fig. 2c), which may be promoter associated small RNAs that were not exported, or short-lived promoter upstream transcripts (PROMPTS).

### Proximity-CLIP reveals features of abundant non-coding RNA classes

We asked whether Proximity-CLIP provided insights into features of nuclear and cytoplasmic lncRNAs and crossed our data with a previously curated list of 61 lncRNAs, of which about 30 were expressed in our system. lncRNAs may serve as miRNA precursors, and our nuclear data set provided evidence of this phenomenon (Suppl. Fig. 3). Several antisense lncRNAs initiated at their annotated TSS, however their transcript signal decays gradually, suggesting that their expression may be regulated during transcription elongation, or that they are products of bidirectional transcription rather than being stable transcripts of their own (Suppl. Fig. 4). Several lncRNAs were found expressed in the RNAseq of total cell extracts, but Proximity-CLIP revealed their transcription initiated upstream of the annotated TSS, in what appeared like divergent transcription from neighboring genes (Suppl. Fig. 5). Some transcripts annotated as lncRNAs appeared to be readthrough transcription products of protein-coding genes (Suppl. Fig. 6). Finally, we also detected intriguing, yet anecdotal expression patterns of lncRNAs: Nuclear and cytoplasmic RAB30-AS1 exon1 was differentially spliced (Suppl. Fig. 7); a highly abundant uncharacterized small RNA, only detectable as an RBP footprint, was transcribed 7kb upstream of TERC (Suppl. Fig. 8); RBP footprints suggested that XIST was dominantly protein-bound at the 5' ends of its first and last exons (Suppl. Fig. 9); and while MALAT1 is highly expressed in HEK293 cells, mascRNA was completely undetectable (Suppl. Fig. 10).

Next, we asked whether we could observe other non-coding RNAs processed in the nucleus and focused on pri-miRNAs that are cleaved by the DROSHA/DGCR8 complex in the nucleus to generate ~80–100 nt long precursor miRNAs (pre-miRNAs) hairpins and exported to the cytoplasm to be further processed. As expected our cytoplasmic RBP

footprints exclusively covered pre-miRNA coordinates, and particularly the sequences corresponding to the mature miRNAs, while nuclear footprints allowed for the detection of the short-lived pri-miRNAs (Fig. 3f, Suppl. Fig. 2d).

To refine our analysis of small RNAs, we divided our footprints prior to cDNA library construction into two sizes, smaller and larger than 40 nt (Suppl. Fig. 2e), reasoning that longer footprints may only stem from unprocessed miRNA precursors, rather than the 21–23 nt long mature species. For the longer footprints cytoplasmic coverage dropped dramatically while nuclear footprints coverage did not change, indicating that we were indeed isolating miRNA precursors (Fig. 3g). Separating long and short footprints may also prove useful to probe mature tRNA molecules and protein-bound tRNA fragments (tRF) (Suppl. Fig. 2f-j). We captured full-length tRNAs with the long footprints (Suppl. Fig. 2f) and were also able to visualize the drop in coverage around the introns found in a number of tRNAs (Suppl. Fig. 2g). Analysis of the short footprints (Suppl. Fig. 2h-j) demonstrated the potential to measure tRFs: Consistent with previous reports from HEK293 cells cytoplasmic coverage at tRF-3 coordinates is higher than at tRF-5 coordinates<sup>38,39</sup> (Suppl. Fig. 2h,j), and we could also detect a signal at tRNA 3' tail coordinates (Suppl. Fig. 2i) that may correspond to the previously reported tRF-1.

RBP footprints on mRNAs could provide insights into RNA regulatory elements occupied by RBPs in different compartments. Among the best characterized and reliably predictable *cis*-acting elements on mRNA are the miRNA binding sites– occupied by the RNA induced silencing complex (RISC). We extracted all predicted miRNA binding sites of conserved miRNAs and calculated their RBP footprint coverage in the cytoplasm and the nucleus. In agreement with known RISC function in most cell types, we found that predicted miRNA binding sites were predominantly occupied in the cytoplasm, but not in the nucleus (Fig. 3h). Furthermore, we asked whether we could detect differences in cytoplasmic miRNA binding site occupancy of sites predicted to be regulated by miR-16, the highest expressed miRNA in HEK293, versus the targets of a non-expressed miRNA, miR-124. As expected, we only found RBP footprint coverage for miR-16 sites, suggesting that Proximity-CLIP was indeed capturing RNA regulatory elements (Fig. 3i). Taken together, our results indicate that Proximity-CLIP is able to identify enriched proteins, RNAs, as well as RNA-protected footprints in our proof-of-principle compartments, the nucleus and the cytoplasm.

### Proximity-CLIP identifies RNAs and proteins enriched at cell-cell interface

Nuclear and cytoplasmic RNPs can be easily fractionated using biochemical methods and thus, we wanted to apply Proximity-CLIP to a compartment that is not accessible to other fractionation approaches. We chose to study the plasma membrane compartment at the cell-cell interface because it may be a site of localized translation, as well as of intercellular communication, which may involve RNA-dependent signaling and regulation. Considering that exogenously expressed CNX43 is incorporated in cell-cell gap-junctions, we generated a stable HEK293 cell line expressing a CNX43-EGFP-APEX2 fusion protein that indeed targeted APEX2 to cell-cell interfaces and enabled compartment-specific, BP- and H<sub>2</sub>O<sub>2</sub>-dependent biotinylation of proteins (Fig. 4a, Suppl. Fig. 11a).



Standardizing CNX-43 Proximity-CLIP by a cytoplasmic Proximity-CLIP control resulted in a list of 612 cell-cell interface enriched proteins, of which we defined 229 as a high-stringency list (Suppl. Data 4). Functional enrichment analysis of the relaxed and stringent protein lists resulted in similar trends revealing annotations related to the cell-cell interface, such as “cell-cell adherens junction” or “Cell junction” (Fig. 4b, Suppl. Fig. 11b). To further validate our approach we compared our data to a proteome related to another cell-cell interface marker - Cadherin - that we defined from available data— obtained using different cell lines and techniques. While the localization of Cadherin and CNX43 only partially overlaps, the two proteomes shared significant similarity (Fig. 4c); 40% of CNX43-proximate proteome was also found to be Cadherin-related (p-value  $2.2 \times 10^{-6}$ ).

We also profiled RBP protected footprints at cell-cell interfaces and found that their distribution across mRNA annotation categories was indistinguishable from cytoplasmic footprints (Figs. 2h, 4d). We defined a set of 400 mRNAs enriched among cell-cell-interface transcripts, as well as a high-stringency list of 19 mRNAs found enriched in all data sets (Suppl. Data 5). mRNAs of cell interface proteins were overrepresented, demonstrating the ability of Proximity-CLIP to isolate from the larger population of cellular transcripts a subset of mRNAs that potentially undergo local translation or regulation. The majority of enriched mRNA however, encoded either proteins related to transcriptional regulation or transcription factors (Fig. 4e). Consistently, among the top 19 localized RNAs, 10 encoded membrane-localized proteins, while 5 encoded nuclear transcription factors, and altogether 16 encode regulatory proteins (Fig. 4f). This result raises the possibility that on top of localized translation, this intricate membrane compartment also harbors mRNAs that are regulated in an external stimulation-dependent manner. Finally, in order to identify *cis*-acting elements that may play a role in RNA localization and regulation at cell-cell interfaces, we performed motif searches that revealed a significant enrichment of a CUG sequence elements in the RBP protected footprints from 3'UTRs of the top 400 mRNAs at cell-cell interfaces (Fig. 4g, Suppl. Fig 11c). Interestingly, a previous study reported the role of CUG-binding proteins in mRNA transport to the plasma membrane, and we were able to exclusively detect one of those RBPs, CUGBP1, in the proteome localized at cell-cell interfaces (Suppl. Data 4).

## Discussion

Precise control of RNA localization and translation enables cells to maintain proper protein distribution<sup>2,6</sup> and substantial effort was invested to elucidate the subcellular localization of various transcripts. Here, we present Proximity-CLIP, a high-throughput method that allows for simultaneous profiling of localized RNA including short-lived species, characterize enriched *cis*-acting elements, as well as candidate interacting RBPs. The method is easy to use and can be adapted to query any subcellular compartment target by a localization signal across multiple cell types. Proximity-CLIP thus holds substantial advantages over previous approaches.

In contrast to imaging-based technologies, Proximity-CLIP offers higher throughput and by simultaneously probing for the proteome of the compartment of interest, Proximity-CLIP offers information on potential interaction partners and the local environment. Finally,

Proximity-CLIP does not require chemical fixation or permeabilization of cells prior to analysis, steps that can increase experimental noise or perturb RNA localization.

RNA and protein localization was previously studied in high-throughput using fractionation-based approaches. However, not all cellular compartments are accessible to fractionation, fractionation schemes may prove experimentally challenging, and incomplete separation can result in contaminations and/or false positive or negative identification of localized RNA. In addition, Proximity-CLIP does not require the preservation of cellular entities, allowing harsh extraction and purification conditions that result in rapid quenching of undesired cellular reactions, e.g. ribonuclease activity, that may distort transcript identification and quantification. Nevertheless, in contrast to fractionation-based approaches that can be performed in unmodified cells and tissues, Proximity-CLIP requires the expression of APEX2 targeted to the compartment of interest and labeling with 4SU and BP practically restricting its use to cultured cells.

Very recently, Ting and colleagues reported a conceptually similar approach to ours, termed APEX-RIP. In contrast to Proximity-CLIP that uses UV-crosslinking, APEX-RIP uses formaldehyde (FA)-crosslinking of RNA and interacting proteins. However, FA is not a zero-distance crosslinker and also stabilizes protein-protein interactions and crosslinking conditions thus need to be carefully optimized to avoid long-distance crosslinking and detection of indirect interactions. Furthermore, our use of 4SU as photoreactive nucleoside results in a diagnostic T-to-C mutation at the site of crosslinking that allows for convenient bioinformatic removal of noise from copurified RNAs that interact unspecifically with the affinity chromatography matrix. Finally, FA crosslinking occurs after quenching of the APEX2 reaction, while Proximity-CLIP UV crosslinking occurs during quenching, significantly reducing the time gap between protein biotinylation and protein-RNA linkage, minimizing diffusion and RNP rearrangements.

A distinctive feature of Proximity-CLIP is the sequencing of RBP protected footprints that not only allows for profiling of localized RNAs, but also for the identification of protein-occupied *cis*-acting elements on RNA. In contrast to previous global mapping of *cis*-acting elements, our approach provides a snapshot of regulatory elements on RNA that are occupied in the examined compartments. A subset of our RBP footprints appear to be derived from short-lived RNA species, likely by crosslinking to their processing enzymes, such as RNA Polymerase or nuclear RNases and the detection and quantification of RNA intermediates might contribute to kinetic studies of RNA processing and turnover.

Application of Proximity-CLIP to multiple membranous compartments will help answer the open question how proteins localize to concrete membrane niches. Localized translation could enable the specific localization of peripheral, lipid- and tail- anchored membrane proteins, and their mRNA would be guided by RBPs to these sites. In addition, localizing mRNAs with their encoded proteins enables tuning of their expression level according to confined needs to promote protein homeostasis of organelles and pathways. Interestingly, on top of mRNAs that encode cell-cell interface proteins we mostly found enriched mRNAs that encode gene regulatory proteins. Although hypothetical, the translation and stability of mRNAs that encode response factors to extracellular signals might benefit from being



locally regulated at sites where extracellular information is sensed, such as the plasma membrane, cell-cell interfaces, or at sites along the endocytic pathway.

## ONLINE METHODS

### Plasmids, cell lines and media

pCDNA3 Connexin43-GFP-APEX2<sup>1</sup> and pCDNA3-APEX2-NES were gifts from Alice Ting (Addgene plasmids # 49385 and 49386, respectively). All enzymatic manipulations for plasmid generation described below were performed using standard reaction conditions according to the enzyme manufacturer's instructions.

To construct pENTR221(APEX2-NES), the APEX2-NES sequence was amplified by PCR using pCDNA3-APEX2-NES (Addgene 49386) as template and primers #5 and #6 (Suppl. Table 1), and introduced by BP ligation into pDONR221 (Invitrogen) according to manufacturer' instructions. To construct pDEST(frt)(N-ter\_FLAGHA, C-ter AgeI site), pDEST(frt)(FLAGHA) was amplified by PCR using primers #11 and #12, and self-ligated using T4 DNA ligase (NEB). To construct pEXP(FLAGHA-APEX2-NES), an LR reaction (Invitrogen) was performed to recombined the Apex2-NES sequence from pENTR221(APEX2-NES) into pDEST(frt)(N-ter\_FLAGHA, C-ter AgeI site). Then, to construct pEXP(V5-APEX2-NES), pEXP(FLAGHA-APEX2-NES) was amplified by PCR using primers #9 and #10, and self-ligated. pEXP(V5-APEX2-NES) is available on Addgene (#107596).

To construct pENTR221(APEX2), the APEX2 sequence was amplified by PCR using pCDNA3-APEX2-NES (Addgene 49386) as template and primers #5 and #22, and introduced by BP ligation into pDONR221 (Invitrogen). To construct pENTR221(H2B-APEX2) an H2B megaprimer was amplified by PCR using pFC15A-H2B (a gift from Gordon Hager, NCI) as template and primers #24 and #25, and introduced by restriction-free (RF) cloning into pENTR221(APEX2). To construct pDEST(frt)(N-ter\_V5, C-ter AgeI site), pDEST(frt)(N-ter\_FLAGHA, C-ter AgeI site) was amplified by PCR using primers #9 and #10, and self-ligated. Then, to construct pEXP(V5-H2B-APEX2), an LR reaction (Invitrogen) was performed to recombine the H2B-APEX sequence from pENTR221(H2B-APEX2) into pDEST(frt)(N-ter\_V5, C-ter AgeI site). pEXP(V5-H2B-APEX2) was submitted to Addgene (#107597).

Stable cell lines inducibly expressing V5\_H2B\_APEX2 or V5-APEX2-NES were prepared according to the manufacturer instructions (Invitrogen) and as previously described: Flp-In T-REx HEK293 cells (Invitrogen) were grown in standard media (DMEM[Gibco, 11995–065], 10% FBS, 1:1000 PenStrep [Gibco, 15140–122]), supplemented with 100 µg/ml Zeocin [frt site selection] + 15 µg/ml Blasticidin [tetR selection] (“pre-selection medium”). Cells were co-transfected with pEXP(V5-APEX2-NES) or pEXP(V5\_H2B\_APEX2) (Addgene plasmids #107596 and #107597 respectively) and pOG44 plasmid (Invitrogen) in the absence of PenStrep, and selected in standard growth medium supplemented with 100 µg/ml Hygromycin [pFRT insert] + 15 µg/ml Blasticidin (“post-selection medium”). To prepare a stable cell line expressing APEX2-EGFP-Connexin43, Flp-In T-REx HEK293

cells were transfected with pCDNA3-Connexin43-GFP-APEX2 (Addgene plasmid # 49385), and selected by standard media supplemented with 350 µg/ml G418.

### Proximity-CLIP

In plating cells for Proximity-CLIP we aimed to achieve high, yet not full, confluence at day of labeling (80–90%), which for our scheme requires at least  $20 \times 10^6$  cells at day of cell splitting to suffice for preparative and control samples, as well as for immunofluorescence (IF). For confocal microscopy, cells were seeded into a 24 well plate over PLL-coated (Sigma, P8920) glass cover slips at a density of  $2.5 \times 10^5$  cells per well. For Western blot analyses,  $2.5 \times 10^6$  cells were seeded per sample into a 6 cm plate. For preparative Proximity-CLIP  $15 \times 10^6$  cells were seeded per sample in 15 cm dishes.

Sixteen hours after seeding 2 µg/ml Doxycycline (DOX) was added to negative control cells (the parental HEK293 T-Rex) and to cell lines inducibly expressing V5-H2B-APEX2 and V5-APEX2-NES. 4-thiouridine (Sigma) (4SU) (100 µM final concentration from a 500 mM stock solution in H<sub>2</sub>O) was added directly to the growth medium in all 6 cm and 15 cm plates. Sixteen hours after DOX induction and 4SU labeling, 500 µM Biotin-Phenol (BP, iris-biotech, Is-3500.1000; using a 500 mM BP in DMSO stock aliquot stored at – 80 °C) was added to the growth medium, and culture plates were returned to the incubator for 30 minutes. During incubation with BP, fresh quenching solution was prepared by combining sodium ascorbate (VWR) (10 mM; 1 M stock solution freshly dissolved in ddH<sub>2</sub>O), Trolox (Sigma, 238813) (5 mM; freshly dissolved 500 mM stock solution in DMSO), and sodium azide (10 mM; 1 M stock solution in water can be stored at –20 °C or below for a year) in pre-chilled PBS (Gibco, 10010–023).

After 30 minutes in the presence of BP, a freshly made stock of 100 mM H<sub>2</sub>O<sub>2</sub> (Sigma) in PBS was added to cell culture medium to a final concentration of 1 mM for 60 seconds, then media was quickly discarded and cells were quickly (but gently, to minimize cell loss) washed three times in large volumes (at least twice the medium volume per wash) of quencher solution. Then, 750 µl of quenching solution was added to stop the biotinylation reaction in plates and 500 µl was added to 24 wells.

Cells growing in 6 or 15 cm plates were subjected to 312 nm UV crosslinking, without plates cap, at 0.15 J/cm<sup>2</sup>, using a Spectrolinker XL-1500 (Spectronics Corporation). Crosslinked cells were collected by gentle scraping and pelleted by centrifugation for 5 min at 300g and 4 °C. The supernatant was discarded cells were snap frozen in liquid N<sub>2</sub>.

### Fluorescence microscopy

Quenched cells in 24-well plates were fixed with 4% PFA in PBS for 20 minutes. Fixed cells were washed 3 times with PBS, and further fixed and permeabilized by adding –20°C cold methanol and further incubation for 20 min at –20°C. Then, cells were blocked with 5% BSA w/v in PBS (“blocking solution”) for 60 min. Labeling with primary mouse-αV5 antibody (R960–25, Thermo scientific) was performed in humid chamber, overnight at 4°C (1:400 dilution in blocking solution), followed by washing 3 times for 5 min with PBS. Then, fixed cells were incubated with Alexa488-coupled Goata.Mouse (Thermo, A11001) and Alexa647-coupled Neuroavidin (self-made, as described in) for 1 hour at room

temperature (1:600 and 1:300 dilutions, respectively). Finally, fixed cells were washed 3 times for 5 min with PBS and mounted over 9  $\mu$ l vectashield with DAPI (VECTOR, H-1200).

### Cell fractionation

Fractionation was performed as described in Gagnon et al., 2014, with a few modifications. Briefly, 10 cells were pelleted in PBS at 250 g and resuspended in 380  $\mu$ l HLB (10 mM Tris, pH 7.5, 50 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% NP-40, and 10% glycerol). After a 2-minute incubation, cells were centrifuged at 400 g and the supernatant (cytoplasm) was obtained. The pellet was washed three times (resuspend in 1 ml HLB and centrifuged at 200 g) with HLB buffer. RNA was extracted from the nuclei using Trizol reagent.

### Cell extraction and Streptavidin pull-down for Western blot controls

Cell pellets originating from 6 cm plates were resuspended in 300  $\mu$ l RIPA buffer (50 mM Tris, 150 mM NaCl, 0.1% (wt/vol) SDS, 0.5% (wt/vol) sodium deoxycholate and 1% (vol/vol) Triton X-100, pH 7.5), supplemented with 1x protease inhibitor cocktail without EDTA (Roche, 04693159001), 1 mM PMSF and fresh quenching solution (10 mM sodium azide, 10 mM sodium ascorbate and 5 mM Trolox, see section 2 above for preparation). The cell suspension was incubated on ice for 2 min, and extracts were clarified by centrifugation at 15,000g for 10 min at 4 °C. Cell extracts were kept on ice throughout the procedure. Protein concentration in cell extracts was quantified using the Pierce 660-nm assay (Pierce, 22660) (typical protein concentration was around 1.2  $\mu$ g/ $\mu$ l).

150  $\mu$ l of RIPA buffer were added to 150  $\mu$ g cell extract for each sample to increase volume and incubated with 15  $\mu$ l of streptavidin magnetic beads pre-washed in RIPA buffer at 4 °C overnight on a rotator. Note: this step can also be done for 1 h at room temperature. Remaining extract was saved for gel and western blot analysis. Notes: When handling the streptavidin magnetic beads, either 1 ml or cut 200  $\mu$ l pipette tips were used. Beads were collected using a magnetic rack and the supernatant (flow-through [FT]) was saved on ice for subsequent analysis. Beads were washed by a series of ice-cold buffers (1 ml for each wash) to remove nonspecific binders, as follows: 2 x with RIPA buffer, 1 x with 1 M KCl, 1 x with 0.1 M Na<sub>2</sub>CO<sub>3</sub>, 1 x with 2 M urea in 10 mM Tris-HCl, pH 8.0 (freshly prepared), and 2 x again with RIPA buffer.

Biotinylated proteins were eluted for WB analysis by heating each sample in 60  $\mu$ l of 3x protein loading buffer (for 6x take 10.5 ml ddH<sub>2</sub>O, 10.5 ml 1 M Tris (pH 6.8), 10.8 ml glycerol, 3 g SDS, 2.79 g DTT and 3.6 mg bromophenol blue) supplemented with 2 mM biotin and additional 20 mM DTT for 10 min (97 °C, while vortexing). Samples were then cooled to room temperature, briefly centrifuged, and placed on a magnetic rack to collect the eluate. For Western blot analysis 10  $\mu$ l of eluate was loaded per lane of a 4–20% (w/v) gradient SDS polyacrylamide gel.

Cell extracts and FT were supplemented with 1x protein loading buffer, heated to 97 °C, chilled, and briefly centrifuged to collect condensate. Samples were loaded at equal protein quantities (accounting for FT volume increase due to addition of RIPA buffer for incubation with streptavidin beads) and run through a 4–20% (w/v) gradient SDS polyacrylamide gel.

### Preparative cell extraction and Streptavidin pull down

Cell pellets originating from 15 cm plates were resuspended in 800  $\mu$ l of RIPA buffer supplemented with 1x protease inhibitor cocktail without EDTA, 1 mM PMSF and fresh quenchers (10 mM sodium azide, 10 mM sodium ascorbate and 5 mM Trolox, preparation, see section 2). Resuspended cells were incubated on ice for 2 min, and extracts were cleared by centrifugation at 15,000g for 10 min at 4 °C. Cell extracts were kept on ice throughout the procedure. Protein concentration in cell extracts was quantified using the Pierce 660-nm assay (Pierce, 22660) (typical protein concentration was around 3  $\mu$ g/ $\mu$ l).

30  $\mu$ l of the cell extracts was saved for total RNA isolation and RNAseq (see below) and another 80  $\mu$ l were saved for protein, gel and western blot analyses. For each sample, 1.5 mg of protein extract (~500  $\mu$ l) was incubated with 60  $\mu$ l of pre-washed streptavidin magnetic beads while rotating at room temperature for 1 h. Beads were pelleted using a magnetic rack and FT was collected and saved for subsequent analysis. Beads were then washed by a series of buffers as described in section 4. At the final wash step the beads were divided into three aliquots: 1.) 300  $\mu$ l for Mass Spectrometry – liquid removed and beads kept on ice. 2.) 200  $\mu$ l for bound-RNA seq (No RNase treatment – liquid was removed and beads were freeze-dried at –80°C). 3.) 500  $\mu$ l for RNase T1 treatment (continued immediately).

### RBP footprinting and radiolabeling of RNA footprints

Beads were resuspended in 100  $\mu$ l of RNase T1 buffer (20 mM Tris pH7.4, 150 mM NaCl, 2 mM EDTA, 1% NP40), supplemented with RNase T1 (Thermo, EN0541) to a final concentration of 1 U/ $\mu$ l and incubated for 15 min at 22 °C. Reactions were chilled on ice for 5 min, and beads were washed twice with RNase T1 buffer and once with dephosphorylation buffer (NEB cutsmart buffer x1). Beads were then resuspended in 60  $\mu$ l of dephosphorylation mix (per 300  $\mu$ l of mix take 30  $\mu$ l of 10x cutSmart, 255  $\mu$ l ddH<sub>2</sub>O, 15  $\mu$ l CIP 10 U/ $\mu$ l), and incubated for 10 min at 37°C with shaking. Note: adjust the shaking speed on the thermomixers so the beads do not settle. Beads were then washed twice with 1 ml of dephosphorylation buffer and twice with 1 ml PNK buffer without DTT (50 mM Tris pH7.5, 50 mM NaCl, 10 mM MgCl<sub>2</sub>). Then, beads were resuspended in 60  $\mu$ l of radioactive reaction mix (per 300  $\mu$ l of mix take - 245  $\mu$ l ddH<sub>2</sub>O, 30  $\mu$ l 10x PNK buffer [NEB], 30  $\mu$ l PNK [NEB] to 1 U/ $\mu$ l, 5  $\mu$ l \*ATP [0.5  $\mu$ Ci  $\gamma$ -<sup>32</sup>P-ATP]) and incubated for 30 min at 37°C with shaking. Note: adjust the shaking speed on the thermomixers so the beads do not settle. Then, non-radioactive ATP was added to a final concentration of 100  $\mu$ M and incubation continued for additional 5 min.

Beads were then washed 5 times with 1 ml of PNK buffer without DTT (exposure to > 1 mM DTT for a prolonged time may damage magnetic beads and should only be used in the reaction buffer). Note: We saved 50  $\mu$ l of the radioactive waste, which is collected after the first bead wash, to mark the Urea-PAGE gels as will be described later. Radioactivity on the beads was estimated using a Geiger counter after completion of washing, and beads were stored at –20°C, while the samples for MS analysis were processed.

## Protein sample preparation and Mass spectrometry

Beads kept on ice (see section 5) were resuspended in 30  $\mu$ l of 25 mM ammonium bicarbonate and 20 mM DTT (Thermo, 20291), and incubated at increasing temperatures for 1 h under constant agitation: 25°C (30 min), 37°C (20 min), 56 °C (10 min). Then, 6  $\mu$ l of 200 mM iodoacetamide (Thermo, 90034) (in 25 mM ammonium bicarbonate) were added and tubes were further shaken for 1 h at 25 °C. Beads were collected on a magnet and washed 3 times in 200  $\mu$ l of 1 mM DTT in 25 mM ammonium bicarbonate, to quench the remaining iodoacetamide and to assure depletion of detergents from previous steps. Freshly dissolved trypsin (Promega, V5111) was dissolved at 20  $\mu$ g/ml in 25 mM ammonium bicarbonate. Then, beads were resuspended in 98  $\mu$ l of 25 mM ammonium bicarbonate, supplemented with 2  $\mu$ l of trypsin (40 ng). Tubes were shaken overnight at 37 °C in a mixer with a lid to avoid condensation. Finally, beads were pelleted and the eluate was collected to fresh tubes for further processing by the proteomics facility. As control, the procedure above was also performed on empty beads.

Mass spectrometry was performed on an Orbitrap Fusion coupled with an Ultimate 3000-nLC (Thermo). Peptides were separated on an EASY-Spray C<sub>18</sub> column (Thermo; 75  $\mu$ m x 25 cm inner diameter, 2  $\mu$ m particle size and 100 Å pore size). Separation was achieved by 5–35% linear gradient of acetonitrile + 0.1% formic acid over 120 min. An electrospray voltage of 2.1 kV was applied to the eluent via the EASY-Spray column electrode. The Orbitrap Fusion was operated in positive ion data-dependent mode. Full scan MS was performed in the Orbitrap with a normal precursor mass range of 350–1500 m/z at a resolution of 120 k. The automatic gain control (AGC) target and maximum accumulation time settings were set to  $4 \times 10^5$  and 50 ms, respectively. MS was triggered by selecting the most intense precursor ions above an intensity threshold of  $5 \times 10^3$  for collision-induced dissociation (CID)-MS fragmentation with an AGC target and maximum accumulation time settings of  $5 \times 10^2$  and 250 ms, respectively. Mass filtering was performed by the quadrupole with a 1.6 m/z transmission window, followed by CID fragmentation in the ion trap (rapid mode) and a normalized collision energy (NCE) of 35%. To improve the spectral acquisition rate, parallelizable time was activated. The number of MS spectra acquired between full scans was restricted to a duty cycle of 3 s.

## Elution of RNA from beads

Beads immobilized RNPs (RNase treated or not, stored at –80 °C and –20 °C, respectively) were proteolytically digested with proteinase K to elute the RNA: Beads were defrosted and the digestion was performed in 3 subsequent steps, each time adding to the existing volume for a final volume of 500  $\mu$ l: 1.) addition of 1.2 mg/ml proteinase K in 200  $\mu$ l of 1x Proteinase K buffer (50 mM Tris pH 7.5, 75 mM NaCl, 6.25 mM EDTA, 1% SDS), followed by incubation at 50°C in a heat block under vigorous shaking for 30 min (for 9 samples - weigh 2.16 mg Proteinase K and resuspend in 1.8 ml buffer). 2.) addition of 0.75 mg/ml proteinase K in 150  $\mu$ l of 1x Proteinase K buffer, followed by incubation at 50°C for 30 min (for 9 samples - weigh 1.01 mg and resuspend in 1.35 ml buffer). 3.) addition of 0.75 mg/ml proteinase K in 150  $\mu$ l of 1x Proteinase K buffer, followed by incubation at 50°C. At the end of the reaction tubes were briefly centrifuged, beads were collected on a magnetic rack and the RNA-containing supernatant was transferred to a new 1.5 ml microcentrifuge tube,

combined with 30  $\mu$ l of 5 M NaCl and 300  $\mu$ l acidic phenol-chloroform (pH 4.5) and mixed by vortexing. After a 10 min incubation at room temperature, tubes were centrifuged at 12,000g for 2 min and the aqueous phase was transferred to a new 1.5 ml microcentrifuge tube, combined with 300  $\mu$ l of chloroform, vortexed and centrifuged at 12000g for 2 min. The aqueous phase was then transferred to a new 1.5 ml microcentrifuge tube, and RNA was precipitated by addition of 1  $\mu$ l of GlycoBlue (Thermo, AM9516) (10 mg/ml), mixing, followed by addition of 3 volumes of ethanol, incubation at  $-20^{\circ}\text{C}$  for at least 1 h, and centrifugation at  $>12,000\text{g}$  and  $4^{\circ}\text{C}$  for 20 min. After removal of all ethanol traces, pellets were air-dried for 5 min at room temperature, and dissolved in 20  $\mu$ l of DEPC-treated ddH<sub>2</sub>O.

### RNA extraction from total cell extracts

RNA from the 30  $\mu$ l samples of total extracts (see section 5) was extracted by consecutive and immediate addition of 370  $\mu$ l DEPC-treated ddH<sub>2</sub>O (to increase the volume of aqueous phase) and 400  $\mu$ l biophenol (Sigma, P3803), vortexing, incubation for 15 min at room temperature and centrifugation for 10 min at max speed. The top 200  $\mu$ l of aqueous phase was transferred to a new tube, added with same volume of ddH<sub>2</sub>O-saturated chloroform, vortexed and centrifuged for 8 minutes at max speed. The top 100  $\mu$ l of aqueous phase were transferred to a new tube, combined with 7  $\mu$ l of 3 M NaAc (pH 5.3) and 400  $\mu$ l of cold EtOH, vortexed and incubated at  $-80^{\circ}\text{C}$  for at least 3 hours. The RNA precipitate was pelleted by centrifugation at  $4^{\circ}\text{C}$  at  $>12,000\text{g}$  for 15 minutes, pellets were washed twice with 500  $\mu$ l of 75% EtOH, air dried for 5 minutes and resuspend in 20  $\mu$ l of DEPC-treated ddH<sub>2</sub>O.

### RNA-seq library prep

RNA-seq libraries were prepared using the NEBnext RNA sequencing kit (E7530) according to the manufacturer's instructions with the following parameters: 1.) rRNA depletion was performed only on total RNA samples using the NEB rRNA depletion kit (E6310). 2.) RNA fragmentation was performed for 10 minutes. Used barcodes are listed in Suppl. Table 1, all fastq files were uploaded to GEO after de-multiplexing.

### sRNA library prep

Small RNA cDNA libraries were prepared as previously described<sup>7,8</sup>: To determine the size of RBP protected RNA fragments samples were loaded and separated by denaturing polyacrylamide electrophoresis (National Diagnostics, EC830/835/840 in 1x TBE), next to RNA size markers (Suppl. Table 1). RNA footprints were visualized by autoradiography and bands representing estimated lengths of 20–40 and 41–70 nt were excised and extracted from gel as follows: 1.) 1 min centrifugation at  $>12,000\text{g}$  in gel breaker tubes (IST Engineering, 3388). 2.) addition of 350  $\mu$ l of 0.3 M NaCl. 3.) shaking for 1 hour at  $60^{\circ}\text{C}$ . 4.) centrifugation for 1 min at 5000g in filter tube (IST Engineering, 5388). 5.) addition of 1  $\mu$ l Glycoblu and 1,200  $\mu$ l EtOH, followed by vortexing, incubation at  $-80^{\circ}\text{C}$  for 20 min, centrifugation for 15 min at  $>12,000\text{g}$  and  $4^{\circ}\text{C}$ . The pellet was air-dried for 5 minutes, and resuspended in 8.7  $\mu$ l DEPC-treated ddH<sub>2</sub>O.



Adapter ligation at the 3' end of footprints was performed by addition of 6  $\mu$ l 50% DMSO, 2  $\mu$ l 10X RNA ligase buffer [wo ATP] (NEB), 0.3  $\mu$ l  $^{32}$ P labeled 19/35 size marker mix (see for preparation instructions), and 2  $\mu$ l of 10  $\mu$ M adenylated 3' adapter (adapter sequence in Suppl. Table 1), followed by incubation for 1 min at 90 °C. The reaction was chilled on ice for 1 min and 1  $\mu$ l T4 Rnl2(1–249)K227Q (1 $\mu$ g/ $\mu$ l) (NEB) was added and the reaction was incubated overnight on ice.

To separate the ligated product from unligated footprints and excess of adenylated adapter samples were combined with 20  $\mu$ l of denaturing PAA gel loading solution, incubated for 1 min at 90 °C, and separated on a 15% denaturing PAA gel in 1x TBE. RNA was visualized by autoradiography and ligated footprints with inserts comparable to the 19 and 35 nt marker (for 20–40b footprints) and above (for 40–70b footprints) were excised (see also Suppl. Fig. 3e). Nucleic acids were extracted from gel as described above and resuspended in 9  $\mu$ l of DEPC-treated ddH<sub>2</sub>O.

Adapter ligation at the 5' end was performed by addition of 1  $\mu$ l of 100  $\mu$ M 5' adapter (Suppl. Table 1), 2  $\mu$ l 10x RNA ligase buffer with ATP (Thermo) and 6  $\mu$ l 50% DMSO, denaturing for 1 min at 90 °C followed by chilling of tubes on ice for 1 min, addition of 2  $\mu$ l Rnl1 (1 mg/ml, Thermo), and a 1 h incubation at 37 °C.

To separate the ligated product from unligated footprints and excess of adenylated adapter samples were added with 20  $\mu$ l of denaturing PAA gel loading solution, incubated for 1 min at 90 °C, and loaded and separated on a 12% denaturing PAA gel. RNA was visualized by autoradiography and ligated footprints were excised as above. Nucleic acids were extracted from gel as described above and resuspended in 4.6  $\mu$ l of DEPC-treated ddH<sub>2</sub>O.

Ligated RNA footprints were denatured by incubation at 90 °C for 1 min, cooled to 50 °C in a thermocycler and reverse transcribed by addition of 1.5  $\mu$ l 100 mM DTT, 3  $\mu$ l 5 $\times$  1st strand buffer (Invitrogen), 4.2  $\mu$ l 2 mM dNTPs, 1  $\mu$ l 100  $\mu$ M reverse PCR primer (3' barcoded primer, Suppl. Table 1), and 0.7  $\mu$ l Superscript III reverse transcriptase (Thermo), followed by incubation for 1 hour at 50 °C. The cDNA product was diluted by addition of 85  $\mu$ l of DEPC-treated ddH<sub>2</sub>O to a volume of 100  $\mu$ l, and 6  $\mu$ l of it were used as template for a diagnostic PCR as follows.

A 60  $\mu$ l diagnostic PCR reaction was set by addition of 38.6  $\mu$ l ddH<sub>2</sub>O, 6  $\mu$ l 10x Platinum Taq buffer without Mg (Invitrogen, 10966018), 1.8  $\mu$ l 50 mM MgCl<sub>2</sub>, 6  $\mu$ l 2 mM dNTPs, 0.6  $\mu$ l 100  $\mu$ M 3' barcoded primer, 0.6  $\mu$ l 100  $\mu$ M 5' primer (Suppl. Table 1), and 0.42  $\mu$ l Platinum-Taq polymerase, and split into 6 tubes, one for each of the following PCR cycle numbers: 9,11,13,15,17,19; using the the following protocol: 2 min initial denaturation at 94 °C, cycling for 45 s at 94 °C, 85 s at 50 °C, 60 s at 72 °C, followed by cooling to 4 °C. PCR products were separated on a 2.5% agarose gel in 1x TBE at 90 V for 1 h, and optimal number of cycles was selected per library. Then an identical PCR mixture was assembled to a volume of 300  $\mu$ l and the PCR performed as above for the optimal number of cycles (split to 3 tubes of 100  $\mu$ l). The PCR product was concentrated using the ZYMO PCR purification kit to a 70  $\mu$ l volume, of which 30  $\mu$ l were further purified to deplete the residual primers

using pipenPrep. Expected sizes - Linker-Linker: 126 bp, Short footprint libraries: 146–166 bp, Long footprints libraries: 166–196 bp.

### Data analysis and statistics

For Mass Spectrometry (MS) data analysis including Label Free Quantification (LFQ), MS raw data files were loaded into the Maxquant software (v1.6.0.1) and analyzed with default parameters, using Uniprot reviewed human proteins annotation (Suppl. Data 6). Proteins only identified by site (a Maxquant definition), and other potential contaminants as defined by Maxquant were filtered out. Also, abundant translation machinery and cytoskeleton proteins were disregarded. Finally, proteins that were detected in control samples (blank runs, Trypsinized void samples and samples from cells not expressing APEX2) were also disregarded. For protein detected in both control and experiment compartment, hits were ranked based on LFQ ratio  $\text{compartment}_{\text{experiment}}/\text{compartment}_{\text{control}}$  and filtered by a razor+unique peptides detection threshold as detailed in Suppl. Files 1 and 4. For proteins that were only detected in the experiment compartment hits were ranked based on their LFQ value and filtered by a razor+unique peptides detection threshold as detailed in Suppl. Data 1 and 4. Stringent and / or relaxed protein lists were curated per experimental replicate according to parameters detailed in Suppl. Data 1 (for nuclear and cytoplasmic proteins) and 4 (for cell-cell interface proteins). E-cadherin related proteins list was curated by combining a literature based collection and two experimental screens<sup>11,12</sup> (see list as a tab in Suppl. Data 4).

Sequencing was performed on an Illumina HiSeq 2500 or 3000 platform. Fastq files were retrieved by bcl2fastq conversion software (Illumina) and de-multiplexed by bcl2fastq conversion software and cutadapt (the raw data was deposited to GEO after de-multiplexing under GSE110380). RNA-seq reads were aligned to the human genome version hg19 using TopHat and reads and RPKM per gene were calculated by RNAcounter ([https://bbcf.epfl.ch/bbcflib/tutorial\\_rnacounter.html](https://bbcf.epfl.ch/bbcflib/tutorial_rnacounter.html)) with default parameters. Genes with name beginning as “Hist”, “hla-” and “MTRNR” were often miscounted by RNAcounter and were removed when relevant. For comparison between Proximity-CLIP and nucleu-cytoplasmic fractionation (Fig. 2) reads per gene by RNAcounter were depleted of miRNAs snoRNAs histons-coding mRNAs. Only cell fractions RNA was rRNA-depleted prior to sequencing. Thus for RPKM calculations total number of reads only regarded reads that mapped within the .GTF file coordinates (and not the entire genome) excluding rRNA and intergenic mapped reads. Plotting and Spearman correlation statistics were performed by R. Partek® software was used to dissect mapped data from fractioned nuclei RNAseq to intronic- versus exonic-mapped reads.

For RBP footprints, sequence reads were mapped to the human genome (hg19) and clusters of overlapping sequence with diagnostic T-to-C mutations identified using the PARalyzer software incorporated into a pipeline (PARpipe; [https://ohlerlab.mdc-berlin.de/software/PARpipe\\_119/](https://ohlerlab.mdc-berlin.de/software/PARpipe_119/)) with default settings. Binding sites were categorized using the Gencode GRCh37.p13 GTF annotation (gencode.v19.chr\_patch\_hapl\_scaff.annotation.gtf,<http://www.gencodegenes.org/releases/19.html>). Metagene analysis of footprints coverage along protein coding transcripts was performed using metaplotR. All other coverage analyses

along genomic coordinates were performed using NGSplot. Visualization of sequence coverage was performed using IGV<sup>16,17</sup>. Conserved miRNA targets coordinates were taken from miRcode (<http://www.mircode.org/download.php>), tRNA-related coordinates were taken from, splicing branch points coordinates were taken from. Sequence motifs were analyzed by ssHMM, and 3' UTR clusters were defined by depletion of clusters that account for "5'UTR", "5'UTR-intron", "coding", "coding-intron", "intron", "intron-5'UTR", "intron-coding", "start\_codon", "snRNA", "snoRNA", "rRNA", "mt\_tRNA", or "mt\_rRNA". Functional enrichment analyses were performed using DAVID<sup>21,22</sup>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank the NHLBI proteomics core and A. Aponte and M. Gucek for Mass Spectrometry performance and analysis, as well as E. Anderson for additional advice on Proteomics data analysis. We would like to thank the NIAMS Genomics Core Facility and G. Gutierrez-Cruz and S. Dell'Orso for sequencing support. We also want to acknowledge the NIH HPC Biowulf cluster, the NIAMS Biodata Mining and Discovery Section and S. Brooks, H.-W. Sun, and D. Heller for computational resources and support. Finally, we want to thank the NIH Medical Arts Branch and A. Hoofring for designing Figure 1. This work was supported by the Intramural Research Program of the National Institute for Arthritis and Musculoskeletal and Skin Disease.

## REFERENCES

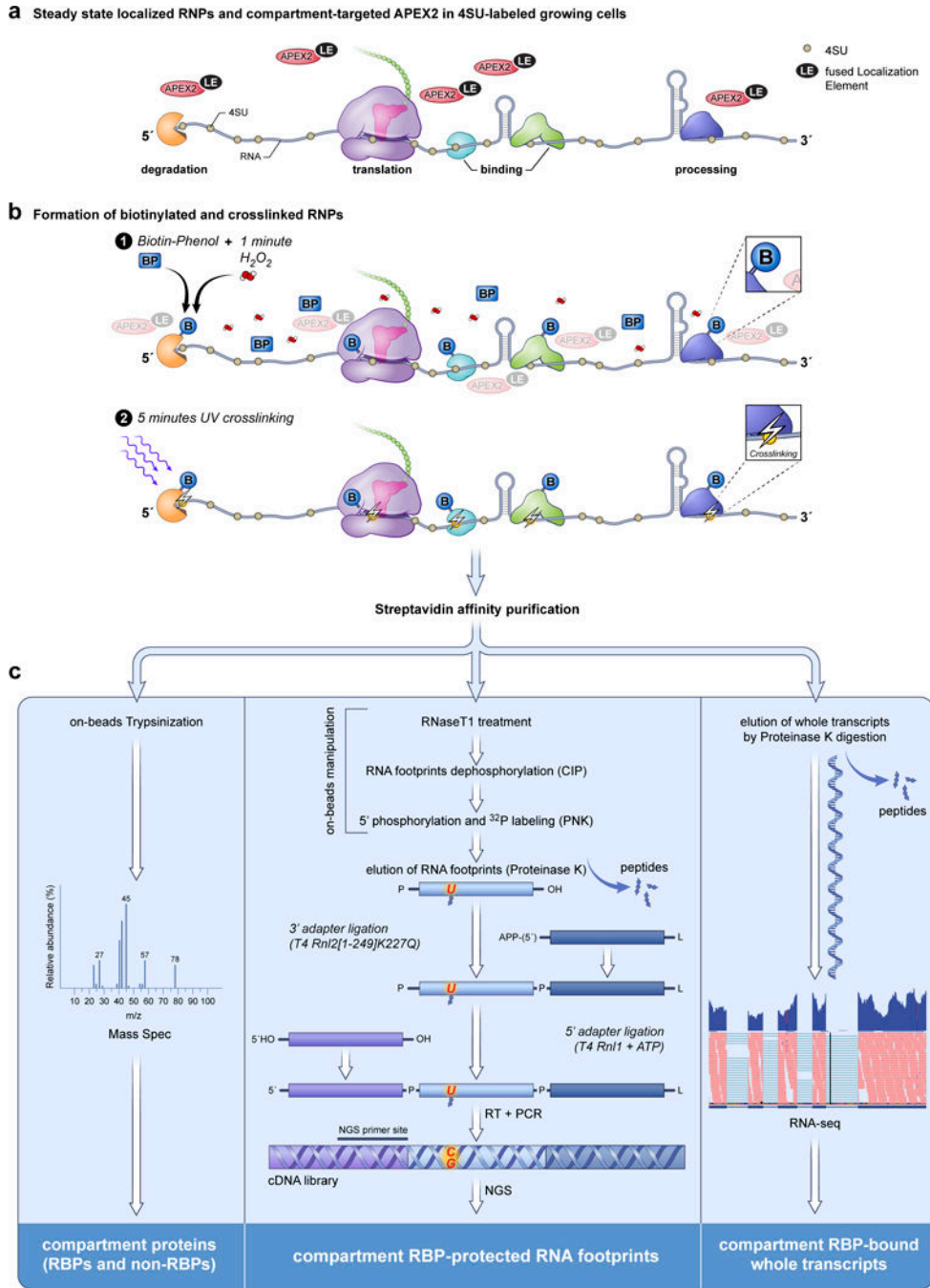
1. Taliaferro JM, Wang ET & Burge CB Genomic analysis of RNA localization. *RNA Biology* 11, 1040–1050 (2014). [PubMed: 25483039]
2. Lécuyer E et al. Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell* 131, 174–187 (2007). [PubMed: 17923096]
3. Ephrussi A, Dickinson LK & Lehmann R oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* 66, 37–50 (1991). [PubMed: 2070417]
4. Long RM et al. Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA. *Science* (80-. ). 277, 383–387 (1997).
5. Wilk R, Hu J, Blotsky D & Krause HM Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. *Genes Dev.* 30, 594–609 (2016). [PubMed: 26944682]
6. Kejiou NS & Palazzo AF mRNA localization as a rheostat to regulate subcellular gene expression. *Wiley Interdisciplinary Reviews: RNA* 8, 1–11 (2017).
7. Hung V et al. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* 11, 456–475 (2016). [PubMed: 26866790]
8. Kaewsapsak P, Shechner DM, Mallard W, Rinn JL & Ting AY Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* 6, e29224 (2017). [PubMed: 29239719]
9. Chen CL & Perrimon N Proximity-dependent labeling methods for proteomic profiling in living cells. *Wiley Interdiscip. Rev. Dev. Biol.* 6, (2017).
10. Hafner M et al. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* 141, 129–141 (2010). [PubMed: 20371350]
11. Choder M mRNA imprinting. *Cell. Logist.* 1, 37–40 (2011). [PubMed: 21686103]
12. Gerstberger S, Hafner M & Tuschl T A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845 (2014). [PubMed: 25365966]
13. Castello A et al. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* 149, 1393–1406 (2012). [PubMed: 22658674]

14. Baltz AG et al. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* 46, 674–690 (2012). [PubMed: 22681889]
15. Lee SY et al. APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. *Cell Rep.* 15, 1837–1847 (2016). [PubMed: 27184847]
16. Huang DW, Sherman BT & Lempicki RA Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57 (2009). [PubMed: 19131956]
17. Slany A et al. Contribution of Human Fibroblasts and Endothelial Cells to the Hallmarks of Inflammation as Determined by Proteome Profiling. *Mol. Cell. Proteomics* 15, 1982–1997 (2016). [PubMed: 27025457]
18. Corcoran DL et al. PARalyzer: Definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* 12, R79 (2011). [PubMed: 21851591]
19. Elkon R, Ugalde AP & Agami R Alternative cleavage and polyadenylation: Extent, regulation and function. *Nature Reviews Genetics* 14, 496–506 (2013).
20. Tian B & Manley JL Alternative cleavage and polyadenylation: The long and short of it. *Trends Biochem. Sci.* 38, 312–320 (2013). [PubMed: 23632313]
21. Schueler M et al. Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol.* 15, (2014).
22. Cenik C et al. Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet.* 7, (2011).
23. Palazzo AF et al. The signal sequence coding region promotes nuclear export of mRNA. *PLoS Biol.* 5, 2862–2874 (2007).
24. Carmody SR & Wente SR mRNA nuclear export at a glance. *J. Cell Sci.* 122, 1933–1937 (2009). [PubMed: 19494120]
25. Mercer TR et al. Genome-wide discovery of human splicing branchpoints. *Genome Res.* 25, 290–303 (2015). [PubMed: 25561518]
26. Valadkhan S & Manley JL Splicing-related catalysis by protein-free snRNAs. *Nature* 413, 701–707 (2001). [PubMed: 11607023]
27. Luo W & Bentley D A ribonucleolytic rat torpedo RNA polymerase II. *Cell* 119, 911–914 (2004). [PubMed: 15620350]
28. Vilborg A, Passarelli MC, Yario TA, Tycowski KT & Steitz JA Widespread Inducible Transcription Downstream of Human Genes. *Mol. Cell* 59, 449–461 (2015). [PubMed: 26190259]
29. Vilborg A et al. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc. Natl. Acad. Sci.* 114, E8362–E8371 (2017). [PubMed: 28928151]
30. Almada AE, Wu X, Kriz AJ, Burge CB & Sharp PA Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360–363 (2013). [PubMed: 23792564]
31. Core LJ & Lis JT Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* 319, 1791–1792 (2008). [PubMed: 18369138]
32. Ntini E et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* 20, 923–928 (2013). [PubMed: 23851456]
33. Seila AC et al. Divergent transcription from active promoters. *Science* (80-. ). 322, 1849–1851 (2008).
34. Fejes-Toth K et al. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature* 457, 1028–1032 (2009). [PubMed: 19169241]
35. Preker P et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* (80-. ). 322, 1851–1854 (2008).
36. Cabili MN et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 16, (2015).
37. Gogakos T et al. Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep.* 20, (2017).
38. Haussecker D et al. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 16, 673–695 (2010). [PubMed: 20181738]

39. Kumar P, Anaya J, Mudunuri SB & Dutta A Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Med.* 12, 1–14 (2014).
40. Xie X et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434, 338–345 (2005). [PubMed: 15735639]
41. Bartel DP Metazoan MicroRNAs. *Cell* 173, 20–51 (2018). [PubMed: 29570994]
42. Agarwal V, Bell GW, Nam JW & Bartel DP Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, 1–38 (2015).
43. Mittelbrunn M & Sánchez-Madrid F Intercellular communication: Diverse structures for exchange of genetic information. *Nature Reviews Molecular Cell Biology* 13, 328–335 (2012). [PubMed: 22510790]
44. Toyofuku T et al. Intercellular calcium signaling via gap junction in connexin-43-transfected cells. *J. Biol. Chem.* 273, 1519–1528 (1998). [PubMed: 9430691]
45. Zaidel-Bar R Cadherin adhesome at a glance. *J. Cell Sci.* 126, 373–378 (2013). [PubMed: 23547085]
46. Guo Z et al. E-cadherin interactome complexity and robustness resolved by quantitative proteomics. *Sci. Signal.* 7, (2014).
47. Van Itallie CM et al. Biotin ligase tagging identifies proteins proximal to E-cadherin, including lipoma preferred partner, a regulator of epithelial cell–cell and cell–substrate adhesion. *J. Cell Sci.* 127, 885–895 (2014). [PubMed: 24338363]
48. Wang ET et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 150, 710–724 (2012). [PubMed: 22901804]
49. Moravcevic K, Oxley CL & Lemmon MA Conditional peripheral membrane proteins: Facing up to limited specificity. *Structure* 20, 15–27 (2012). [PubMed: 22193136]
50. Heller D, Krestel R, Ohler U, Vingron M & Marsico A ssHMM: extracting intuitive sequence-structure motifs from high-throughput RNA-binding protein data. *Nucleic Acids Res.* 45, 11004–11018 (2017). [PubMed: 28977546]
51. Lam, S. S et al. Directed evolution of Apex2 for electron microscopy and proximity labeling. *Nat. Methods* 12, 51–54 (2015). [PubMed: 25419960]
52. Landthaler M et al. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* 14, 2580–2596 (2008). [PubMed: 18978028]
53. van den ent F & LÖWE JRF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. BioPhys. Methods* 67, 67–74 (2006). [PubMed: 16480772]
54. Spitzer J, Landthaler M & Tuschl T Rapid creation of stable mammalian cell lines for regulated expression of proteins using the Gateway® recombination cloning technology and Flp-In T-REx® lines. *Meth. Enzymol.* 529, 99–124 (2013). [PubMed: 24011039]
55. Hung V et al. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat Protoc* 11, 456–475 (2016). [PubMed: 26866790]
56. Gagnon KT, Li L, Janowski BA & COREY DR Analysis of nuclear rna interference in human cells by subcellular fractionation and Argonaute loading. *Nat Protoc* 9, 2045–2060 (2014). [PubMed: 25079428]
57. Benhalevy D, McFarland HL, Sarshad AA & Hafner M PAR-CLIP and streamlined small RNA cDNA library preparation protocol for the identification of RNA binding protein target sites. *Methods* 118–119 41–49 (2016). [PubMed: 27871973]
58. Hafner M et al. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods* 58, 164–170 (2012). [PubMed: 22885844]
59. Cox J & MANN M Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 2011 29:1 26, 1367–1372 (2008).
60. Zaidel-Bar R Cadherin adhesome at a glance. *J. Cell. Sci.* 126, 373–378 (2013). [PubMed: 23547085]

61. Van Itallie CM et al. Biotin ligase tagging identifies proteins proximal to E-cadherin, including lipoma preferred partner, a regulator of epithelial cell-cell and cell-substrate adhesion. *J. Cell. Sci.* 127, 885–895 (2014). [PubMed: 24338363]
62. Guo Z et al. E-cadherin interactome complexity and robustness resolved by quantitative proteomics. *Sci Signal* 7, RS7–RS7 (2014). [PubMed: 25468996]
63. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, PP. 10–12 (2011).
64. Olarerin-George AO & Jaffrey SR MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. *Bioinformatics* 33, 1563–1564 (2017). [PubMed: 28158328]
65. Shen L, Shao N, Liu X & Nestler E ngs.Plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15, 284 (2014). [PubMed: 24735413]
66. Robinson JT et al. Integrative genomics viewer. *Nature Biotechnology* 2011 29:1 29, 24–26 (2011).
67. Thorvaldsdóttir H, Robinson JT & Mesirov JP Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178–192 (2013). [PubMed: 22517427]
68. Gogakos T et al. Characterizing Expression and Processing Of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell REP* 20, 1463–1475 (2017). [PubMed: 28793268]
69. Mercer TR et al. Genome-wide discovery of human splicing branchpoints. *Genome Res.* 25, 290–303 (2015). [PubMed: 25561518]
70. Heller D, Krestel R, Ohler U, Vingron M & Marsico A ssHMM: extracting intuitive sequence-structure motifs from high-throughput RNA-binding protein data. *Nucleic Acids Res.* 45, 11004–11018 (2017). [PubMed: 28977546]
71. Huang DW, Sherman BT & Lempicki RA Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44–57 (2009). [PubMed: 19131956]
72. Huang DW, Sherman BT & Lempicki RA Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13 (2009). [PubMed: 19033363]

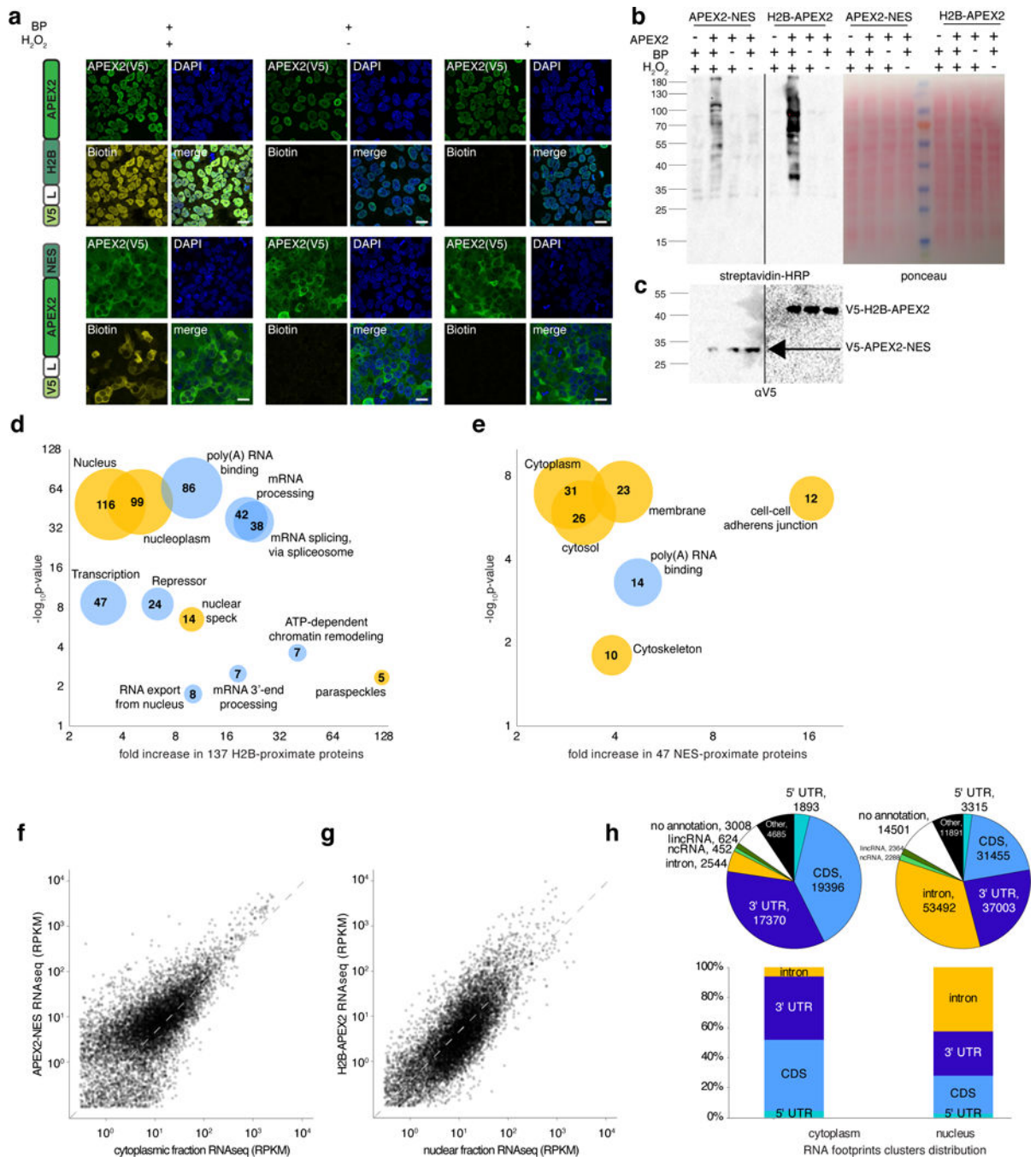




**Figure 1: Proximity-CLIP scheme**

| **a**, Proximity-CLIP takes advantage of the occupancy by RBPs of cellular RNAs throughout their life cycle. An APEX2 fusion protein is targeted to a cellular compartment of choice using a fused localization element (LE), and cellular RNAs are labeled with 4SU. **b**, Cells are incubated with biotin-phenol (BP) for 30 min, before APEX2-mediated BP oxidation is activated by addition of hydrogen peroxide, followed by reaction quenching and 4SU-dependent protein-RNA crosslinking by UV. BP radicals are created locally and either covalently tag proximate proteins or decay. Compartment-specific RNPs and proteins are

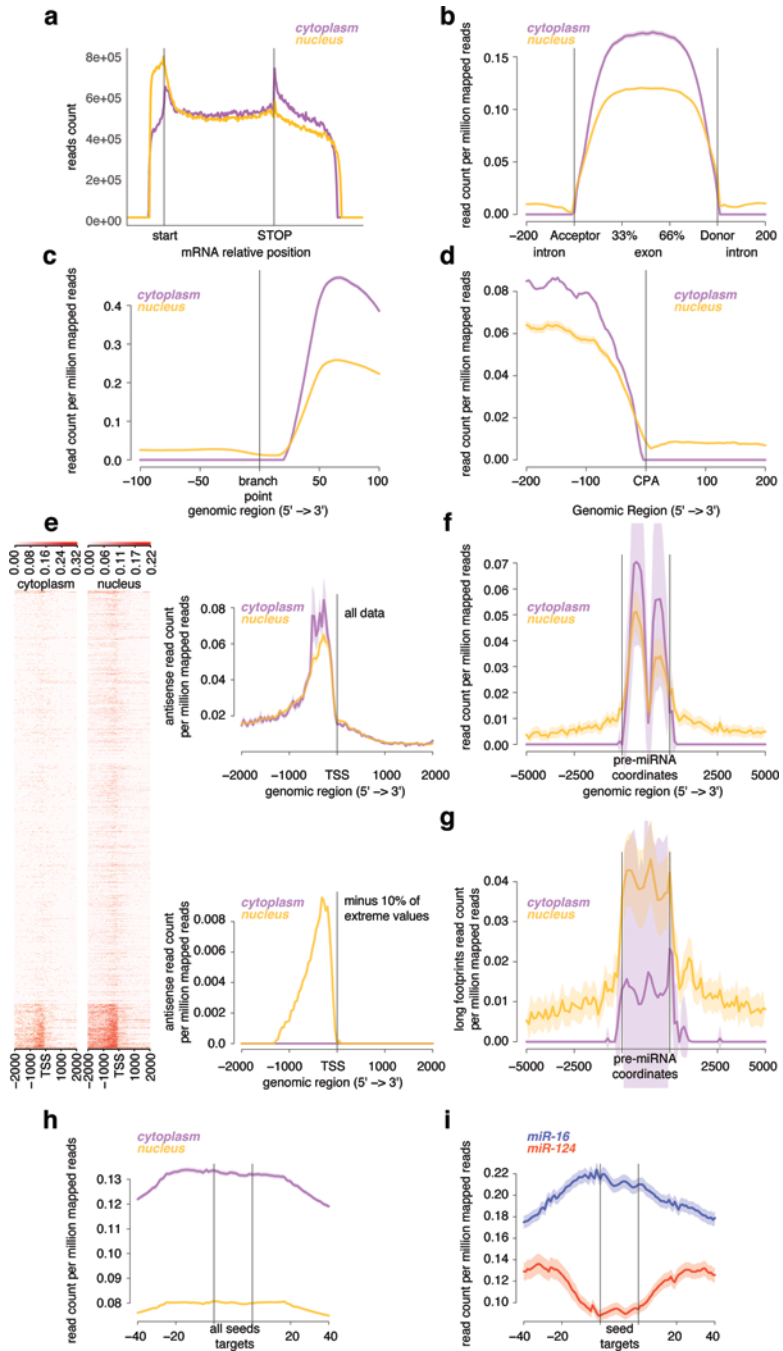
captured by streptavidin affinity chromatography. **c**, The eluate from **b** is split in three parts: The compartment proteome is analyzed by mass spectrometry (left panel) and the RNA is either treated by RNase and analyzed by small RNA cDNA library preparation of RBP-protected footprints, analogous to PAR-CLIP (middle panel), or by standard RNAseq (right panel).



**Figure 2: Proximity-CLIP accurately identifies proteins and transcripts localized to the nucleus or cytoplasm**

**a**, Left, V5-tagged APEX2 constructs inducibly expressed by stable HEK293 cell lines. Right, cells were incubated with either or both H<sub>2</sub>O<sub>2</sub> and BP. V5, DAPI, biotin and merged channels were obtained by immunofluorescence confocal microscopy, scale = 20 μm, scanned fields, >30, documented fields, 5. **b**, Extracts of 4SU-labeled and crosslinked cells described in **a** were analyzed by Western blot; left, Streptavidin-HRP, right, Ponceau S stain of the nitrocellulose membrane. **c**, Anti-V5 western blot analysis of cell extract from **b**.

Results are representative of two independent experiments with identical results. **d**, Functional enrichment analysis of nuclear protein hits list, devised by stringent analysis. Enrichment statistics were obtained by DAVID, with the FDR p-value multiple hypothesis correction. **e**, As **d** for cytoplasmic protein hits. **f**, Reads Per Kilobase per million Mapped reads (RPKM) obtained by APEX2-NES Proximity-CLIP versus cytoplasmic biochemical fractionation. **g**, As **f**, for H2B-APEX2 Proximity-CLIP versus nuclear biochemical fractionation. **h**, Top, number of protein-occupied sites on all transcripts across annotation categories in APEX2-NES (left) and H2B-APEX2 (right) Proximity-CLIP. Bottom, percentage of protein-occupied sites on mRNA 5' and 3' untranslated regions (UTR), as well as coding (CDS) and intronic sequences.



**Figure 3: Analysis of RBP footprints captures short-lived RNA elements and RNA regulatory elements**

| **a**, Coverage of Nuclear and cytoplasmic RBP-protected footprints along the coordinates of mature mRNAs. **b,c**, Coverage of nuclear and cytoplasmic RBP-protected footprints along exons ± 200 bp (**b**), and relative to the splicing branch point (**c**). **d**, RBP footprint coverage relative to the Cleavage and Poly-Adenylation site (CPA). **e**, Antisense reads coverage at genomic coordinates around the transcription start site (TSS): Heatmap of coverage around all TSS (left). Antisense reads coverage around TSS of 100% of genes (right top), and after

removal of divergent genes (right bottom). **f,g**, Nuclear and cytoplasmic Proximity-CLIPs coverage of 20–40 nt (f) and of 40–70 nt (g) long footprints around pre-miRNAs genomic coordinates. **h,i**, Putative RISC footprints on target mRNAs: (h) Nuclear and cytoplasmic footprints coverage around genomic coordinates of conserved miRNA binding sites. (i), Cytoplasmic RBP footprints around predicted target sites of miRs –16 (highly expressed) and –124 (not expressed).

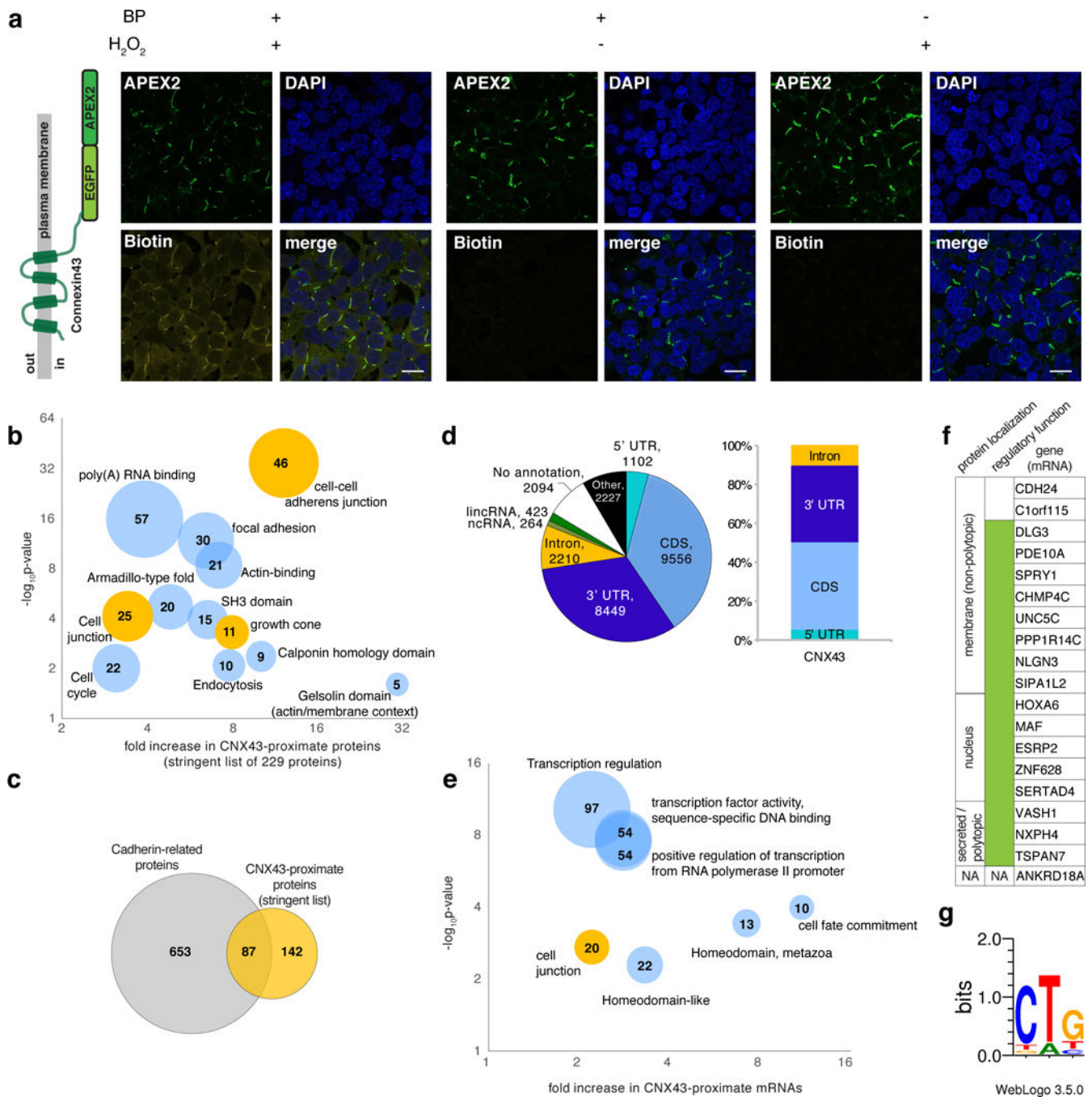
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 4: Proximity-CLIP accurately identifies proteins and transcripts localized to cell-cell junctions**

**a**, Left, topological model of CNX43-EGFP-APEX2 construct expressed by stable HEK293 cell line. Right, cells were incubated with either or both biotin-phenol (BP) and hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>). EGFP, DAPI and biotin channel were obtained by confocal fluorescent microscopy, scale = 20 μm, scanned fields, >30, documented fields, 5. Results are representative of two independent experiments with identical results. **b**, Functional enrichment analysis of cell-cell interface protein hits list, devised by stringent analysis.

Enrichment statistics were obtained by DAVID, with the FDR p-value multiple hypothesis correction. **c**, Venn diagram crossing Cadherin-related proteome based on previous studies—and the protein hits list analyzed in panel **b**. **d**, Left, number of protein-occupied sites on all transcripts across annotation categories. Right, percentage of protein-occupied sites on mRNA 5' and 3' untranslated regions (UTR), as well as coding (CDS) and intronic sequences. **e**, Functional enrichment analysis, as in **b**, of the top 400 enriched mRNAs at cell-cell interfaces. **f**, High stringency list of 19 mRNAs enriched at cell-cell interfaces identified by Proximity-CLIP. For each mRNA, known gene regulatory function is marked in green, and the localization of their encoded proteins is listed. **g**, Weblogo of 3' UTR footprints RNA recognition element enriched at cell-cell interfaces, generated by ssHMM.