



Published in final edited form as:

Nat Methods. 2021 May ; 18(5): 507–519. doi:10.1038/s41592-021-01128-0.

Robust single-cell discovery of RNA targets of RNA binding proteins and ribosomes

Kristopher W. Brannan^{1,2,3}, Isaac A. Chaim^{1,2,3}, Ryan J. Marina^{1,2,3}, Brian A. Yee^{1,2,3}, Eric R. Kofman^{1,2,3}, Daniel A. Lorenz^{1,2,3}, Pratibha Jagannatha^{1,2,3}, Kevin D. Dong^{1,2,3}, Assael A. Madrigal^{1,2,3}, Jason G. Underwood⁴, Gene W. Yeo^{1,2,3,*}

¹Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California, USA

²Stem Cell Program, University of California San Diego, La Jolla, California, USA

³Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA

⁴Pacific Biosciences (PacBio) of California, Inc., Menlo Park, California, USA

Abstract

RNA binding proteins (RBPs) are critical regulators of gene expression and RNA processing that are required for gene function. Yet, the dynamics of RBP regulation in single cells is unknown. To address this gap in understanding, we developed STAMP (Surveying Targets by APOBEC Mediated Profiling), which efficiently detects RBP-RNA interactions. STAMP does not rely on UV-crosslinking or immunoprecipitation and, when coupled with single-cell capture, can identify RBP- and cell type-specific RNA-protein interactions for multiple RBPs and cell types in single, pooled experiments. Pairing STAMP with long-read sequencing yields RBP target sites in an isoform-specific manner. Finally, Ribo-STAMP leverages small ribosomal subunits to measure

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#terms

*Corresponding author geneyeo@ucsd.edu.

Contributions

Conceptualization K.W.B. and G.W.Y.; Methodology, K.W.B. and G.W.Y.; Investigation, K.W.B., I.A.C., K.D.D., A.A.M., D.A.L., and R.J.M.; Formal Analysis, E.K., P.J., K.W.B., B.A.Y., I.A.C., D.A.L. and R.J.M; Writing – Original Draft, K.W.B. and G.W.Y.; Writing – Review & Editing, G.W.Y., K.W.B., I.A.C., B.A.Y., D.A.L., and R.J.M; Funding Acquisition, G.W.Y.; Supervision, G.W.Y.

Declaration of conflicts of interests

GWY is co-founder, member of the Board of Directors, on the SAB, equity holder, and paid consultant for Locana and Eclipse BioInnovations. GWY is a visiting professor at the National University of Singapore. GWY's interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. The authors declare no other competing financial interests.

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gene W. Yeo (geneyeo@ucsd.edu). Important plasmids described in this study will be deposited in the Addgene plasmid repository and available under a standard MTA.

Editor's summary

Surveying Targets by APOBEC Mediated Profiling (STAMP) identifies binding sites of RBPs by C-to-U RNA editing. STAMP is isoform-specific, can be multiplexed, and enables detection of ribosome association in single cells.

Editor recognition statement: Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reviewer recognition statement: Nature Methods thanks Alfredo Castello and the other, anonymous reviewers, for their contribution to the peer review of this work.

transcriptome-wide ribosome association in single cells. STAMP enables the study of RBP-RNA interactomes and translational landscapes with unprecedented cellular resolution.

Introduction

RNA-binding proteins (RBPs) interact with RNA molecules from synthesis to decay to affect their metabolism, localization, stability and translation [1, 2]. Methods for transcriptome-wide detection of RBP-RNA interactions provide insights into how RBPs control gene expression programs and how RNA processing is disrupted in disease states [3, 4]. Immunoprecipitation-based technologies coupled to high throughput sequencing such as RNA Immunoprecipitation (RIP) and Crosslinking Immunoprecipitation (CLIP) are commonly used to identify RBP targets and binding sites across the transcriptome [5]. While RIP-seq is useful for identifying gene targets of an RBP, CLIP-seq can resolve binding sites within different regions of a given target gene, which lends insight into binding functionality, and allows for the discovery of sequence motifs recognized by specific RBPs [6, 7]. The eukaryotic ribosome is itself composed of a collection of RBPs that can interact directly with mRNA coding sequences [3, 8]. Ribosome profiling methods such as ribo-seq have become a mainstay in the evaluation of transcriptome-scale ribosome occupancy [9, 10]. Unfortunately, CLIP and ribosome profiling experimental protocols are labor-intensive, usually require sizable amounts of input material and transcript fragmentation [11, 12], prohibiting single-cell and long-read platform applications. While there has been rapid progress in single-cell measurements of chromatin accessibility [13], gene expression [14, 15], and surface protein-levels [16, 17], there is currently no available technology for measuring RBP- and ribosome-mRNA interactions at single-cell or isoform-aware resolution.

Recent studies have circumvented the need for immunoprecipitation for detecting RBP-RNA interactions by utilizing fusions of RNA-editing or RNA-modifying modules to RBPs of interest to label RNA targets [18–22]. Internal RNA target labeling has been accomplished by the Target of RNA-binding proteins Identified by Editing (TRIBE) approach, which fuses RBPs of interest to the deaminase domain from the ADAR family of RNA-editing enzymes to mark target RNAs with A-to-I edits [18, 20, 21, 23, 24]. These ADAR-mediated approaches have been used to obtain RBP-targets from low input material, but are limited by the sparsity of double-stranded regions proximal to RBP binding sites that are required for ADAR-mediated adenosine to inosine editing [25].

APOBEC1 is a cytosine deaminase that catalyzes RNA cytosine-to-uracil (C-to-U) conversion on single-stranded RNA substrates [26]. Recently APOBEC1 was fused to the m6A-binding YTH domain to identify m6A modification sites on RNAs genome-wide [27]. This approach, termed Deamination Adjacent to RNA modification Target (DART-seq), identified YTH-domain recognized m6A modifications on single stranded mRNAs. However, it was unclear if APOBEC1 fusions would work with general classes of RBPs or even ribosomes. We reasoned that fusion of APOBEC1 to such full-length RBPs could generalize robust, immunoprecipitation-free identification of RBP targets across functional RNA-interaction categories using extremely low or even single-cell input. Here, we

demonstrate the power of such an approach for detecting RBP-RNA targets at the single-cell and single-molecule level. We developed an integrated experimental and computational framework termed STAMP (Surveying Targets by APOBEC Mediated Profiling) which extends the DART-seq approach to demonstrate the discovery of RBP-RNA sites at isoform-specific and single-cell resolution, the deconvolution of targets for multiplexed RBPs, and the cell-type specific binding of an RBP in a heterogeneous mixture of cell-types. Furthermore, by applying STAMP with specific ribosome subunits, we extend this approach for single-cell detection of ribosome association while simultaneously measuring gene expression.

Results

STAMP identifies RBP binding sites without immunoprecipitation

Our strategy for immunoprecipitation-free detection of RBP targets involves fusing full-length RBPs of interest to the cytidine deaminase enzyme APOBEC1, which is known to catalyze C-to-U editing on single-stranded RNA targets (Figure 1A). Upon expression of an RBP-APOBEC1 fusion protein (RBP-STAMP), RBPs direct the deaminase module to their RNA targets leading to C-to-U base conversion proximal to RBP binding sites. These mutations (edits) are resolved using high-throughput RNA sequencing approaches and quantified using the SAILOR analysis pipeline [28], which we modified to identify and assign a confidence value for C-to-U mismatches using a beta distribution that factors both site coverage and editing percentage following removal of annotated SNPs [29] (see Materials and Methods).

To determine the utility of the STAMP approach, we fused APOBEC1 to the C-terminus of the RBP RBFOX2 [30–32] and generated stable HEK293T cell lines using lentiviral integration. RBFOX2-STAMP is doxycycline inducible to allow modulation of the duration and magnitude of fusion expression, and we noted no detectable change in cell viability or proliferation rate at any induction level or time point. Cells expressing low (50ng/ml doxycycline) and higher (1µg/ml doxycycline) levels of RBFOX2-STAMP for 72 hours had enriched C-to-U edit clusters on the 3' untranslated region (3'UTR) of the known RBFOX2 target *APP* mRNA, and these edit clusters coincided with reproducible RBFOX2 binding sites as detected by enhanced CLIP (eCLIP) of either endogenous RBFOX2 [33] or the RBFOX2-APOBEC1 fusion (Figure 1B). Uninduced RBFOX2-STAMP, or control-STAMP (APOBEC1 only) at low and high induction, had few or no detectable C-to-U edits in the same region, indicating target specificity. RBFOX2-STAMP induced edits within this *APP* 3'UTR target region were 10-fold to 25-fold more frequent than background control-STAMP edits at 0.9 and 0.999 (SAILOR) confidence thresholds, respectively (Figure 1C, Table S1, Table S2). These results demonstrate that fusion of the APOBEC1 module to a well characterized RBP enriches for target-specific edits.

To evaluate the reproducibility of STAMP, we conducted replicate control- and RBFOX2-STAMP with low and high doxycycline inductions for 24, 48 and 72 hours. The number of edited reads (E) on each target gene, normalized to read depth and gene length (EPKM, Edited-reads Per Kilobase of transcript, per Million mapped reads), were highly reproducible and correlations between replicates improved substantially upon induction

($R^2=0.32$ at no dox treatment, to $R^2=0.72$ and 0.83 at low and high dox, respectively; Figure 1D). Irreproducible discovery rate (IDR) analysis [34] also revealed reproducible windows with edits for RBFOX2-STAMP, and the number of these reproducible edits also increased with dox induction of RBFOX2-STAMP (Figure S1A). We also evaluated the effects of RBFOX2-STAMP editing on target transcript levels by conducting differential gene expression analysis on low and high-induction RBFOX2-STAMP at multiple timepoints and detected negligible changes in cellular gene expression compared to uninduced controls (Figure S1B features the results for the 72-hour timepoint, which is similar to the 24 hour and 48 hour results; Table S3). We observed expected basal leakiness of the doxycycline system but with induction, RBFOX2-APOBEC1 mRNA levels increase to within 1.5-fold of the endogenous RBFOX2 levels (Figure 1E).

We next measured the nucleotide distance of RBFOX2-STAMP edits from the conserved RBFOX2 binding site motif. For 2,852 RBFOX2 eCLIP peaks that harbor the canonical RBFOX2 motif UGCAUG, distances from the motif to RBFOX2-STAMP and control-STAMP (background) edits were determined within a 400 bp window (Figure 1F). We observed enriched edits for RBFOX2-STAMP within 200 bp of binding site motifs inside eCLIP peaks, compared to edits from control-STAMP, and the proximity of edits to motifs correlated with eCLIP peak fold enrichment over size-matched input control, indicating that RBFOX2 RNA-binding activity is directing and enriching RBFOX2-STAMP specific edits at conserved sites.

Next, we developed a set of criteria that retrieves high-confidence edit-clusters for RBP-STAMP while reducing false positives, analogous to peak-calling in analyzing CLIP-seq datasets. We observed that the overlap of RBFOX2-STAMP edits with RBFOX2-APOBEC1 eCLIP peaks increases with increasing gene expression thresholds (Figure S1C), and we also anticipated more background edits within more highly expressed substrates. To minimize this background while enriching for true binding sites, we developed an edit cluster-finding algorithm with gene-specific thresholds that assume Poisson-distributed edit-scores e calculated for each site (described in Materials and Methods). Sites that satisfied gene-specific e thresholds ($p < 0.05$ with adjusted Bonferroni correction for multiple-hypothesis testing) and SAILOR confidence score thresholds were then merged with neighboring sites. Instances of edit sites with no neighboring edits within 100 bases in either direction were removed (workflow schematized in Figure S1D). These criteria established a set of 5044 edit-clusters for RBFOX2-STAMP (5.4% of the original unfiltered windows) and removed essentially all background control-STAMP sites (21 remaining, 0.04% of unfiltered windows) (Figure S1E). Next, we determined the fraction of RBFOX2-APOBEC1 eCLIP peaks detected by these RBFOX2-STAMP edit-clusters. We found that nearly half of all significant eCLIP peaks (4-fold enriched over size-matched input and $p < 0.001$) overlapped with RBFOX2 edit-clusters at a SAILOR confidence threshold of 0.9 for the edit sites, more than two-fold higher compared to overlaps with size-matched randomly shuffled regions on exons of the same target genes (Figure 1G). At higher SAILOR confidence thresholds, the fraction that overlaps decreases but the enrichment over background is preserved. We observed that 47% of RBFOX2-STAMP edit-clusters overlapped with RBFOX2 eCLIP peaks, irrespective of whether the eCLIP peaks contained known RBFOX2 binding motifs and an additional 8% of the edit-clusters contained the RBFOX2 motif

(Figure 1H). Interestingly, most clusters that did not overlap with eCLIP peaks were nevertheless located within eCLIP target genes at a distance from neighboring eCLIP peaks (Figure S1F). Subjecting control-STAMP sites to our same criteria for RBFOX2-STAMP sites left essentially no background edit clusters to compare to eCLIP (Figure S1E). We also evaluated the orientation of the APOBEC1 fusion protein and observed that edit-clusters and eCLIP peaks overlapped substantially from APOBEC1 N-terminally fused to RBFOX2. However, the overlap was 20% smaller than what was observed for the C-terminal fusion, demonstrating that fusion orientation should be considered for each RBP of interest to maximize binding site capture (Figure S1G). Lastly, we performed *de novo* motif discovery using high-confidence RBFOX2-STAMP edit-clusters, assessing enrichment above a shuffled background for each gene region. These edit-clusters were significantly enriched for the UGCAUG RBFOX2 binding motif, and the enrichments were correlated with the doxycycline dose and subsequent expression levels of RBFOX2-STAMP (Figure 1I) demonstrating the sensitivity and specificity of STAMP for discovering RBP binding sites.

We next generated two additional HEK293T RBP-STAMP cell lines, one that inducibly expresses APOBEC1 fused to the histone stem-loop binding protein SLBP, and another that expresses a fusion to the stress granule protein TIA1 that binds target mRNA 3'UTRs [3, 35, 36]. We noted similar STAMP-fusion expression levels compared to endogenous TIA1 and SLBP, as we observed for RBFOX2-STAMP (Figure S1H). As with RBFOX2-STAMP, we saw that the number of TIA1-STAMP edits on target genes increased with doxycycline concentration and were strongly correlated across replicates, with summary IDR analysis revealing thousands of reproducible edits that increased in number with increasing induction levels (Figure S1I). Comparison of SLBP-STAMP to SLBP-APOBEC1 eCLIP data showed that SLBP-STAMP edits were enriched compared to control-STAMP near eCLIP peaks within the 3'UTR of histone genes, such as *H2AC16* (Figure 1J) adjacent to stem loop regions, as expected. Comparison of control- and TIA1-STAMP to TIA1-APOBEC1 eCLIP revealed that there was inducible TIA1-STAMP edit enrichment overlapping the example eCLIP 3'UTR peak within the *NPM1* gene (Figure 1K). Globally we found greater than 70% of all significantly reproduced SLBP eCLIP peaks [37] (> 4-fold enriched over size-matched input and $p < 0.001$, reproducible by IDR) overlapped with SLBP-STAMP edit-clusters (Figure S1J), and more than 30% of all significant TIA1-APOBEC1 eCLIP peaks by the same criteria overlapped with TIA1-STAMP edit-clusters (Figure S1K), with size-matched randomly shuffled regions on exons of the respective genes showing significantly lower concordance with edit-clusters at any threshold for both RBPs. We also obtained from *de novo* motif analysis the known eCLIP established U(A)-rich binding sequence from TIA1-STAMP edit-clusters (Figure S1L). These results confirm the versatility of the STAMP approach in specifically and reproducibly detecting the targets and binding sites of multiple RBPs.

Ribosome-subunit STAMP (Ribo-STAMP) edits are enriched in highly translated coding sequences and responsive to mTOR inhibition.

Since ribosomes have extensive association with mRNAs during translation, we reasoned that ribosomal subunits fused to APOBEC1 (Ribo-STAMP) have the potential to edit mRNAs in a manner that reflects ribosome association. We generated independent

HEK293T cell lines expressing APOBEC1 fusions to ribosomal subunits RPS2 and RPS3. For RPS2-STAMP and RPS3-STAMP we observed that edits were enriched relative to control-STAMP on exons of protein-coding genes that are highly translated in HEK293T cells, such as *ATP5PB* [38], coincident with RPS3 eCLIP signal enrichment over size-matched input control (Figure 2A). In comparison, RPS2-STAMP and RPS3-STAMP signal were minimally detected on highly expressed non-coding genes such as the lncRNA *MALATI*, which is localized to the cytoplasm in mitotic cell lines [39](Figure 2B). We performed replicate RPS2-STAMP and control-STAMP inductions at low and high doxycycline concentrations for 24, 48 and 72 hours and again observed dose-dependent STAMP-fusion expression compared to endogenous RPS2 levels (Figure S2A) with strong EPKM reproducibility between replicates ($R^2=0.6$ to 0.8) as well as low overlap (2.8% of all detectable edits) between control-STAMP and RPS2-STAMP edit sites at high induction (Figure S2B–D, Table S4). As edits from RPS2- and RPS3-STAMP were present in coding sequences (CDS) and also in 3'UTR sequences (Figure 2A), we needed to determine if these 3'UTR edits should be filtered or if they are coincident bystander edits. Comparison of EPKM computed from CDS only, to EPKM values computed from both CDS and 3'UTR revealed a strong correlation, indicating that 3'UTR edits need not be excluded from downstream analyses and in some instances may provide edits otherwise missed if we considered only CDS regions in genes with short open-reading frames ($R^2=0.78$, Figure S2E).

To evaluate whether Ribo-STAMP can distinguish genes with varying levels of ribosome-occupancy we next compared combined genome-wide EPKM values from control-, RPS2- and RPS3-STAMP to RPKM values from ribosome protected fragments (RPF) obtained from standard ribosome profiling (ribo-seq) [40] and to RPKM values from poly-ribosome-fraction-enriched RNA (polysome-seq) [41] experiments performed in HEK293 cells. For control-STAMP and for uninduced RPS2-STAMP, EPKM values were poorly correlated with ribo-seq RPKM values ($R^2=0.32$ and $R^2=0.29$ respectively, Figure 2C, **left**). At low and high levels of doxycycline induction, we found that the correlations between EPKM values for RPS2-STAMP and ribo-seq RPF RPKM values improved significantly ($R^2=0.41$ and $R^2=0.46$ respectively Figure 2C). We observed a similar relationship when comparing RPS3-STAMP to ribo-seq ($R^2=0.42$ Figure 2D). RPS2-STAMP and polysome-seq measurements were also well correlated ($R^2=0.54$), consistent across replicates and improved at higher doxycycline induction concentrations and expression times (Figure 2E, Figure S2F). As RPS2-STAMP had higher correlation with independent ribosome foot-printing approaches than RPS3-STAMP (Figure 2D), we proceeded with RPS2-STAMP as the representative Ribo-STAMP fusion for downstream analysis. Meta-coding gene analysis of RPS2-STAMP edits for the top quartile of ribosome-occupied (ribo-seq) genes revealed enrichment of edits within the CDS when compared to control-STAMP background edits and RBFOX2-STAMP edits, which showed the expected 3'UTR profile consistent with eCLIP (Figure 2F). Enrichment of RPS2-STAMP edits within 3'UTRs likely indicates small ribosomal subunit association with these accessible regions following ribosome translation termination by release factors, as we also observed 3'UTR signal from endogenous RPS3 eCLIP (Figure 2A). These results are in agreement with previous studies revealing widespread 3'UTR

ribosome footprints in both yeast and human cells [42–44]. Together these results demonstrate that Ribo-STAMP edit read-counts track ribosome occupancy measurements.

To determine if Ribo-STAMP edits detect translational perturbations, we performed stable high-induction RPS2- and control-STAMP and simultaneously treated cells with the mammalian target of rapamycin (mTOR) pathway inhibitor Torin-1, a selective ATP-competitive inhibitor of mTOR kinase [45]. Pharmacological inhibition of the mTOR pathway globally suppresses translation of mRNAs after initially suppressing translation of genes encoding the translational machinery itself [46]. 72-hour Torin-1 treatment resulted in reproducible suppression in RPS2-STAMP edit distributions compared to vehicle treated cells, exemplified by a marked decrease in edits on the top quartile of ribosome occupied genes (ribo-seq, Figure 2G). RPS2-STAMP EPKM values were also significantly reduced upon Torin-1 treatment in the highest quartile of ribosome occupied genes as defined by ribo-seq (Q1 $p = 1.9 \times 10^{-147}$, Wilcoxon rank-sum test), and polysome-seq (Q1 $p = 7.7 \times 10^{-108}$, Wilcoxon rank-sum test), and all previously reported Torin-1 sensitive TOP genes [46] were contained within these top quartiles (Figure 2H). We observed no significant reduction in EPKM values for control-STAMP cells upon Torin-1 treatment for any matched comparisons (Figure S2G). Gene-level comparison of EPKM values for Torin-1 and vehicle treated RPS2-STAMP on the highest quartile of ribosome occupied genes as defined by ribo-seq revealed translation suppression from Torin-1 treatment (Figure S2H), with no corresponding difference in RPKM values between treated and untreated samples (Figure S2I). Together these results demonstrate that dynamic translational responses are detected by Ribo-STAMP.

Long-read STAMP reveals isoform-specific binding profiles.

To determine if STAMP enables RNA target detection on full-length mRNA isoforms using long-read sequencing technology, we performed 72-hour stable high-induction RBFOX2- and control-STAMP and directly sequenced cDNA long reads with the Oxford Nanopore Technologies (ONT) and PacBio (PB) sequencing platforms [47–50]. Both long-read sequencing approaches resulted in edit enrichment above control from RBFOX2-STAMP that overlapped with both eCLIP signal and short read (Illumina) RBFOX2-STAMP signal, as illustrated by the target gene *APP* 3' UTR (Figure 3A). Long-read high confidence (0.99) RBFOX2-STAMP edits enriched the known RBFOX2 UGCAUG binding motif for both approaches (Figure 3B). As PacBio has a lower base-calling error rate than Nanopore sequencing [51], we observed clear separation between control-STAMP and RBFOX2-STAMP edits and more significant motif extraction. We therefore focused on long read data from PacBio sequencing for downstream isoform-specific editing analysis.

To evaluate isoform-specific binding events, we calculated RBFOX2-STAMP or control-STAMP edit read fractions on the primary and secondary alternative polyadenylation APA isoforms of all genes (RBFOX2-STAMP $n = 1604$, control-STAMP $n = 1878$) that satisfied a minimal coverage threshold of 10 reads per isoform for long reads obtained from PacBio sequencing. We observed differential isoform editing signatures for RBFOX2-STAMP compared to control-STAMP (Figure S3A, Table S5). To illustrate, we displayed edits on the *FAR1* (Figure 3C, D) as well as *PIGN* (Figure 3C, Figure S3B) genes and observed

RBFOX2-STAMP (but not control-STAMP) APA isoform-specific 3'UTR edits, suggesting that RBFOX2 interacts with one of the isoforms but not the other. These isoform-specific binding sites coincided with both short-read RBFOX2-STAMP edit clusters and RBFOX2-APOBEC1 eCLIP peaks, however the association of RBFOX2 to either isoform was indiscernible using these short-read approaches. These results demonstrate that STAMP enables isoform-aware long-read detection of RBP-RNA interactions.

Detection of RBFOX2-RNA targets at single-cell resolution

To evaluate whether STAMP can discover RBP-RNA interactions in single cells, we modified our plasmid vectors to enable capture by the 10x Genomics Single Cell 3' v3 beads and performed 72-hour stable high-induction RBFOX2- and control-STAMP in distinct HEK293T cell-lines followed by standard single-cell (sc)RNA-seq. Using the inserted capture-sequence adjacent to the RBP open-reading frames to identify “capture cells” we identified 844 RBFOX2-STAMP cells and 5,242 control-STAMP cells.

Comparison of bulk and single-cell edit fractions for control- and RBFOX2-STAMP experiments across the top 200 expressed genes (ranked by transcripts per million from bulk RBFOX2-STAMP RNA-seq) revealed nearly identical edit enrichment profiles of RBFOX2 samples above controls and further uncovered a spectrum of editing frequencies across individual cells (Figure 4A). To illustrate, we next ranked individual control- and RBFOX2-STAMP cells by summed ϵ score and visualized edit fractions for the top 10 cells on the RBFOX2 eCLIP target gene *UQCRH*. For all 10 selected RBFOX2-STAMP cells, but not control-STAMP cells, we saw consistent edit signal in close proximity to the RBFOX2 eCLIP peak that overlapped with edit enrichment from both bulk RBFOX2-STAMP and the aggregate of all RBFOX2-STAMP cells (Figure 4B), revealing that STAMP can define RBP binding sites at single-cell resolution. We saw very strong concordance (80%) in the target genes that contained filtered high-confidence RBFOX2-STAMP edit clusters obtained from single-cell and bulk datasets (Figure S4A). At the binding site level, 60% of these high-confidence single-cell edit-clusters directly overlapped edit-clusters obtained from bulk RBFOX2-STAMP, and ~70% of all single cell edit-clusters fell within 400 bp of bulk edit-clusters (Figure 4C). In addition, we found that 73% of single-cell STAMP targets that contained edit-clusters also contained significant RBFOX2-APOBEC1 eCLIP peaks ($P < 0.001$; Figure 4D). As with bulk RBFOX2-STAMP, a majority of the single-cell RBFOX2-STAMP edit-clusters overlapped eCLIP peaks and harbored RBFOX2 binding motifs (Figure 4E), with a large number of clusters that did not directly overlap eCLIP peaks still present in target genes generally within 1000 bp of the neighboring eCLIP peak (Figure 4F). As expected, single-cell RBFOX2-STAMP eCLIP peak capture rate was associated with target expression level (Figure S4B). *De novo* motif analysis from edit-clusters by randomly down-sampling the numbers of single cells analyzed identified the canonical (U)GCAUG motif with significance, even to the resolution of one cell (Figure 4G) showcasing the strength of single-cell STAMP.

Deconvolution of RBP- and cell type-specific RNA binding

The ability of STAMP to recover RBP-RNA targets in single cells suggests that targets of multiple RBPs can be simultaneously discovered from a single multiplexed experiment. In

our RBFOX2-STAMP experiment, we separately performed 72-hour high-induction TIA1-STAMP, prior to mixing equal number of RBFOX2- and TIA1-STAMP cells, followed by scRNA-seq. Cells harboring capture sequences for TIA1- and RBFOX2-STAMP were better distinguished from each other and from control-STAMP cells by UMAP visualization using ϵ scores, than by gene expression (Figure 5A, 5B and S5A), congruent with our expectations that the single-cell ϵ score profiles of TIA1- and RBFOX2-STAMP targets were sufficiently distinct. UMAP visualization of ϵ scores further revealed that control-STAMP cells (n = 8,117 cells) were distinct from RBFOX2- and TIA1-STAMP “capture cells” (Figure 5B). Using Louvain clustering by ϵ score profiles we thus defined an RBFOX2-population (n = 6,003 cells), a TIA1-population (n = 1,841 cells) and a background-population (n = 6,623 cells) for further analysis (Figure S5B). Overlap with control (Figure 5C) and re-clustering in the expression space (Figure S5C) for these defined clusters highlighted the utility of ϵ score-based clustering for defining RBP-specific cell groups. *De novo* motif analysis of edits from the aggregated cells in the RBFOX2-cluster, but not control, confirmed edit enrichment at RBP-specific binding sites (Figure S5D), and TIA1- and RBFOX2-clusters displayed distinct editing profiles when compared to control-STAMP (Figure 5D, Table S5). We ranked cells based on summed ϵ scores to select cells with the most robust editing and found that the top 5 cells for each RBP displayed edit enrichment on the shared RBFOX2- and TIA1-STAMP target *NPM1*, which was also detected as a TIA1 target by eCLIP and bulk TIA1-STAMP (Figure 1K). Edit enrichments for individual cells were specific to TIA1-STAMP on the *BTF3* target gene, and to RBFOX2-STAMP on the *CFL1* target gene (Figure 5E), demonstrating that the targets and binding sites of multiplexed RBP-STAMP fusions can be delineated from edit signatures within single-cell experiments.

To identify cell-type specific RBP targets using single-cell STAMP, we performed STAMP in HEK293T cells and pluripotent stem cell-derived neural progenitor cells (NPCs) [52] by transient transfection with plasmids constitutively expressing either RBFOX2- or control-STAMP fusions, and then mixed equal numbers of HEK293T and NPC cells for each STAMP construct before performing scRNA-seq. UMAP visualization revealed that cells clustered by gene expression into distinct HEK293T and NPC subgroups expressing cell-type specific markers (Figure 5F, Figure S5E, Table S6). UMAP clustering on ϵ score also resulted in separation of cell types (as determined by gene expression clustering) based on RBFOX2-STAMP edits (Figure 5G, Table S7), and we extracted the RBFOX2 binding motif using edit clusters from 2,178 NPC cells editing 468 target genes, and 3,258 HEK293 cells editing 939 target genes (Figure S5F). Analysis of the top RBFOX2-STAMP differentially edited genes between cell types revealed cell-type specific targets (Figure 5H) that were often not differentially expressed (Figure 5I), indicating cell-type specific RNA protein interactions independent of target expression levels. Individual cell edits for the top 5 control- or RBFOX2-STAMP cells from each cell-type ranked by summed ϵ score illustrated targets that were edited specifically in HEK293 cells such as *RPL14* or in NPC cells such as *RPL13A* (Figure 5J). Together, these results indicate that cell type-specific targets and binding sites can be extracted from RBFOX2-STAMP edit signatures by scRNA-seq within a mixture of heterogeneous cell types.

Ribo-STAMP reveals translational landscapes at single-cell resolution.

To examine whether Ribo-STAMP can quantify ribosome association at the single-cell level, we performed stable 72-hour high-induction control- and RPS2-STAMP and conducted scRNA-seq. To distinguish control- and RPS2-STAMP cell populations we computed EPKM measurements for protein-coding genes for each cell. EPKM-based UMAP representation (Figure 6A) followed by Louvain clustering (Figure 6B) revealed a group of RPS2-STAMP (RPS2-population) cells that was clearly distinct from a population of background cells that contained a mixture of both control-STAMP and RPS2-STAMP cells (background-population). Focusing on this RPS2-population, we showed that EPKM values (from CDS and 3'UTR) aggregated from the 3,917 single cells correlated meaningfully ($R^2=0.53$) with genome-wide EPKM values from bulk Ribo-STAMP (Figure S6A). We note that EPKM values computed from edits within the combination of CDS and 3'UTR, versus only CDS regions, correlated very strongly ($R^2=0.81$; Figure S6B), therefore we included 3'UTR-derived edit measurements. We next addressed if aggregated single-cell Ribo-STAMP EPKM values can approximate ribosome-occupancy measurements derived from bulk poly-ribosome-fraction-enriched RNA (polysome-seq). We first assessed if RNA abundance measurements for total mRNA from Ribo-STAMP and polysome-seq experiments were in good agreement and observed a positive relationship ($R^2=0.54$, Figure S6C). We then compared Ribo-STAMP mRNA edits to polysome-seq mRNA abundance and observed less agreement between these measurements ($R^2=0.32$, Figure S6D), suggesting that Ribo-STAMP edit enrichment is not simply dictated by transcript levels. In contrast, poly-ribosome-fraction-enriched RNA measurements from polysome-seq were well correlated with Ribo-STAMP edits ($R^2=0.51$, Figure 6C), implying that single-cell Ribo-STAMP edit enrichments are more closely associated with ribosome occupancy than with transcript abundance. These results strongly indicate that single-cell Ribo-STAMP, like single-cell RBP-STAMP, recapitulates results from bulk experiments and correlates well with standard measurements from orthogonal bulk approaches.

Inspired by our ability to define RBP-specific populations from RBP-STAMP cell mixtures using editing information alone (Figures 5B and 5C), we next integrated Ribo-STAMP with RBP-STAMP to define ribosome association and RBP binding sites in parallel after merging all control-, RBFOX2-, TIA1 and RPS2-STAMP single-cell edits matrices. UMAP visualization of single-cell, transcriptome-wide ϵ scores revealed that control-STAMP cells overlapped with a subpopulation of RBFOX2-, TIA1- and RPS2-STAMP cells (Figure S6E), highlighting cells that have similar background-level edit patterns. Louvain clustering within UMAP projection space defined four distinct groups of single cells for downstream analysis: RPS2-population cells ($n = 3,621$ cells), RBFOX2-population cells ($n = 7,000$ cells), containing the majority (92%) of RBFOX2 cells identified by capture sequencing, TIA1-population cells ($n = 1312$ cells) containing the majority (57%) of TIA1 capture cells, and a background-population ($n = 20,655$ cells), composed of control-STAMP cells and any cells that overlap spatially with control-STAMP cells (Figure 6D, Figure S6F). The ϵ score-derived RPS2-population was 90% matched to the EPKM-derived RPS2-population (Figure S6G), and also had good EPKM value correlation with polysome-seq measurements (Figure 6E). Metagene plotting of edits from these four subgroups for the top quartile of ribosome occupied genes (ribo-seq, $n = 4,931$ genes) demonstrated CDS enrichment for single-cell

RPS2-STAMP edits compared to more 3'UTR-centric enrichment for single-cell RBFOX2- and TIA1-STAMP (Figure 6F), in agreement with our results from bulk (Figure 2F). Differential ϵ score analysis showed distinct editing signatures for RPS2-, RBFOX2- and TIA1-population cells compared to the background-population (Figure 6G, Table S8). To illustrate, the top 10 cells ranked by summed ϵ score exhibited the expected specific editing signatures on the *RPL12*, *RPL30* and *RPL23A* target transcripts (Figure 6H). These results highlight the capacity of STAMP to reveal RBP targets and ribosome association in parallel at single-cell resolution.

Discussion

We have developed an experimental and computational workflow called STAMP (Surveying Targets by APOBEC Mediated Profiling) that allows antibody-free detection of RBP and ribosome interactomes (Ribo-STAMP) by standard RNA sequencing and quantification of binding-site-specific C-to-U edits directed by RBP- and ribosomal subunit-APOBEC1 fusions, respectively. To distinguish our STAMP framework from DART-seq (which uses a portion of the YTH-domain) and TRIBE (which uses ADAR deaminase domains), we showcase unprecedented single-cell resolution binding sites for a range of RBPs and ribosome subunits. Indeed, we were able to demonstrate the specificity of STAMP for full-length RBPs that bind both polyadenylated mRNAs (RBFOX2, TIA1) and non-polyadenylated mRNAs (SLBP). We also demonstrate that the ribosomal subunits RPS2 and RPS3 when fused to APOBEC1 enable the measurement of ribosome association that correlates well with ribosome-occupancy computed from ribo-seq and polysome-seq experiments. In a single experiment, Ribo-STAMP uses edited and non-edited reads to reflect ribosome-associated and input gene expression values simultaneously. We found that Ribo-STAMP signal was sensitive to mTOR pathway inhibition, showcasing responsiveness to specific translational perturbations. We envision that these simultaneous readouts will be extremely useful in more complex and heterogeneous cellular or *in vivo* models to address questions concerning cell identity or disease states. To enable dissemination of our single-cell STAMP technologies, we also developed computational methods that demultiplex multiple RBPs by clustering cells using only edit signatures, which we can validate using 10x feature barcoding technology.

STAMP has distinct advantages over TRIBE, as TRIBE generally yields only gene-level target information and not binding sites, with one to two edits on average detectable in any given target [18, 21, 23, 24]. The sparse editing signal by ADAR deaminase domains is due to the preference for ADAR to edit double-stranded RNAs that contain a bulged mismatch [25], an infrequent occurrence on single stranded mRNAs transcriptome-wide [21, 25, 53]. In contrast, APOBEC enzymes access cytosines in single-stranded RNA which constitute ~25%–35% of nucleotides in any given mammalian transcript and produce clusters of edits (between 10 and 1000 edits at target sites). In addition, structured RNA is reduced from coding regions by active translation [54], making ribosome interactions that are easily detectable by Ribo-STAMP not feasible with ADAR-fusion approaches. Indeed, the RBP FMR1 fusions to ADAR, which is expected to be very frequent in the coding regions of genes resulted in only 4 confident edits across the ~15kb coding region of the showcased *POE* gene [21]. The higher likelihood of encountering APOBEC1 cytosine substrates within

single-stranded mRNA enables STAMP-mediated discovery of RBP-RNA sites with such high sensitivity and specificity that *de novo* discovery of conserved binding-site motifs can be extracted from even one single cell.

Antibody-based methodologies such as CLIP and RIP are staples used to identify RNA binding sites and targets of RBPs. Our STAMP approach offers several advantages. First, CLIP is generally constrained by input requirements, frequently needing thousands to millions of cells. Here we demonstrate that STAMP can be used reliably at single-cell resolution to identify RNA targets, binding sites and even extract motifs from a few to a single cell. STAMP enables the combined identification of RBP binding sites and global measurement of gene expression, a long-standing goal for the gene expression, genomics and RNA communities. Second, CLIP requires fragmentation to separate bound and unbound RNA, but that precludes the discovery of isoform-dependent binding sites on mRNAs that may differ by an exon or translated regions. We show here that STAMP allows long-read assessment to distinguish RBP binding on different transcript isoforms. Further, direct RNA sequencing has recently been demonstrated to be RNA-modification sensitive [55], which opens the possibility of using STAMP to detect modification-sensitive RNA-protein interactions.

In our study, we utilized polyA⁺ mRNA-sequencing (other than total RNA-seq for SLBP-STAMP) to characterize binding interactions for RBFOX2, TIA1, RPS2 and RPS3. However, aside from their mRNA binding functionality, RBFOX2 and TIA1 are also splicing factors that bind intronic regions not detected by polyA selection of mRNAs. Adaptation of the approach to use nuclear isolation, non-polyA selection with the removal of ribosomal RNA contaminants (as we performed for SLBP), or targeted sequencing of intronic regions are strategies anticipated to recover these binding events. False positive binding sites are also possible when expression levels of STAMP transgenes are supra-physiological, leading to promiscuous RNA interactions. For example, while we found that a majority of RBFOX2-STAMP clusters overlapped eCLIP peaks and targets, we did note a number of potential off-target genes containing these clusters, and further study will be necessary to determine if these represent transient physiological RBP-RNA interactions or off-target edits. Alternative approaches to optimize expression for future studies may include use of a native promoter by knocking in the APOBEC deaminase domain in frame on one target cell allele, or transient transfection of synthetic mRNAs that code for the fusion for immediate translation in the cytoplasm.

Currently antibody-free methods like STAMP and TRIBE require fusion of the protein of interest to a modifying enzyme, which may not be feasible for all RBPs. In addition, the editing on this timeframe may have unintended consequences depending on the protein of interest. Our current version of Ribo-STAMP yields detectable edits within 12–24 hours, a timeframe which may dampen the capacity to detect rapid translational responses and may lead to unintended expression modulations due to recoding of transcripts and the possible introduction of nonsense or frameshift mutations. Therefore, it is important to consider the duration of expression of Ribo-STAMP as it relates to the dynamics being assessed and to downstream unintended perturbations. Extended Ribo-STAMP expression could also explain the somewhat unexpected 3'UTR edit enrichment that we observed, although 3'UTR

enrichment of edits appears to be a generalized phenomenon for both TRIBE/HyperTRIBE and DART-seq approaches, likely due to editing modules accessing susceptible 3'UTR sequence elements distal to actual fusion binding sites [20, 21, 23, 27]. Although we demonstrate that this 3'UTR Ribo-STAMP editing is inconsequential to ribosome-occupancy correlations with other profiling methods, these edits may nonetheless be biologically relevant. Gold-standard ribosome profiling (ribo-seq) with alternative digestion conditions in human cells has uncovered widespread 3'UTR ribosome footprinting [44], and translation complex profiling (TCP-seq) in both yeast [43] and human cells [42] revealed small ribosomal subunit specific enrichment within 3'UTRs, likely attributable to ribosomal recycling mediated mRNA interactions. In the future, we anticipate that engineering of fusion orientation, in addition to fine-tuning STAMP expression levels and duration of overexpression window will be useful to obtain editing profiles that are maximally informative for different RBPs and ribosomal subunits. For each new use case of the current version of the STAMP approach, we recommend testing edit signature responses to both fusion orientation and fusion expression level using short-read RNA sequencing from bulk cells and comparing to gold standard orthogonal methods such as CLIP and ribo-seq before proceeding to long-read and single-cell resolution applications.

Looking ahead, as STAMP allows isoform-aware and single-cell level interrogation of RNA-protein interactions, we anticipate that focused genomic integrations of editing modules in animal and organoid models will be powerful for *in vivo* tracing of RNA-protein interaction landscapes in many previously inaccessible contexts. Such model systems expressing STAMP fusions for RBPs of interest hold the potential to unveil the isoform-specific RNA binding and translation landscapes at the organismal level, which would also allow for tissue and cell type-specific profiling in developmental or disease relevant phenotypes.

Materials and Methods

Plasmid construction

For the generation of stable cell lines, all RBP-STAMP mammalian expression constructs were in one of two lentiviral Gateway (Invitrogen) destination vector backbones: 1) pLIX403_APOBEC_HA_P2A_mRuby or 2) pLIX403_Capture1_APOBEC_HA_P2A_mRuby. pLIX403_APOBEC_HA_P2A_mRuby was cloned by amplification (Cloneamp, Takara Bio) of APOBEC1_HA_P2A cassette after removal of the YTH cassette from APOBEC1-YTH (gift from Kate Meyer) originally cloned from pCMV-BE1 plasmid (a gift from D. Liu; Addgene plasmid no. 73019). APOBEC_HA_P2A was inserted into the pLIX403 inducible lentiviral expression vector adapted from pLIX_403 (deposited by David Root, Addgene plasmid # 41395) to contain TRE-gateway-mRuby and PGK-puro-2A-rtTA upstream of mRuby by Gibson assembly reaction of PCR products (Cloneamp, Takara Bio). pLIX403_Capture1_APOBEC_HA_P2A_mRuby was constructed by insertion of a synthetic gene block (Integrated DNA technologies, IDT) containing 10x Feature Barcode Capture Sequence 1 with Gibson assembly reaction into MluI digested backbone pLIX403_APOBEC_HA_P2A_mRuby in frame and immediately upstream of the APOBEC1 ORF. RBP open reading frames (ORFs) were obtained from human Orfeome 8.1

(2016 release) donor plasmids (pDONR223) when available, or amplified (Cloneamp, Takara Bio) from cDNA obtained by SuperScript III (Invitrogen) RT-PCR of HEK293XT cell purified RNA (Direct-zol, Zymogen) and inserted into pDONR223 by Gateway BP Clonase II reactions (Invitrogen). Donor ORFs were inserted in frame upstream of APOBEC1 or Capture Sequence 1 APOBEC1 by gateway LR Clonase II reactions (Invitrogen). For transient transfections of HEK293T cells and NPC cells, constructs were modified from pCMV BE1-YTH-HA plasmid (a gift from Kate Meyer modified from D. Liu; Addgene plasmid no. 73019; <http://n2t.net/addgene:73019>) by removal (control-STAMP) or replacement (RBFOX2-STAMP) of YTH cassette with RBFOX2 open reading frame by PCR and Gibson assembly reactions.

Human cell culture conditions and maintenance

All stable STAMP cell lines were generated using human lenti-X HEK293T cells (HEK293XT, Takara Bio) which are derived from transformed female human embryonic kidney tissue. Cells were maintained in DMEM (4.5 g/L D-glucose) supplemented with 10% FBS (Gibco) at 37° C with 5% CO₂. Cells were periodically passaged once at 70–90% confluency by dissociating with TrypLE Express Enzyme (Gibco) at a ratio of 1:10. The stable HEK293XT cell lines RBFOX2-STAMP, TIA1-STAMP, SLBP-STAMP, RPS2-STAMP, RPS3-STAMP, and control-STAMP were generated as described in Generation of STAMP stable cell lines section. by transducing ~1 million cells with 8µg/ml polybrene and 1ml viral supernatant in DMEM+10%FBS at 37C for 24 hours, followed by subsequent puromycin resistance selection (2µg/ml). Small molecule neural progenitor cells (smNPCs) were grown in medium consisting of DMEM/F12+Glutamax, 1:200 N2 supplement, 1:100 B27 supplement, penicillin/streptomycin (Life technologies), 100mM ascorbic acid (Sigma, A4544), 3mM CHIR99021 (CHIR, Tocris 4423) and 0.5mM Purmorphamine (PMA) (Tocris 4551) and passaged using Accutase. Generation of smNPCs from iPSCs is described in [1].

Generation STAMP stable cell lines

Lentivirus was packaged using HEK293XT cells seeded approximately 24 hours prior to transfection at 30–40% in antibiotic-free DMEM and incubation at 37°C, 5% CO₂ to 70–90% confluency. One hour prior to transfection DMEM was replaced with OptiMEM media transfection was performed with Lipofectamine 2000 and Plus reagent according to manufacturer's recommendations at a 4:2:3 proportion of lentiviral vector: pMD.2g: psPAX2 packaging plasmids. 6 hours following transfection, media was replaced with fresh DMEM + 10% FBS. 48 hours after media replacement, virus containing media was filtered through a 0.45 µm low protein binding membrane. Filtered viral supernatant was then used directly for line generation by transducing ~1 million cells (1 well of 6 well dish) with 8µg/ml polybrene and 1ml viral supernatant in DMEM+10%FBS at 37C for 24 hours. After 24 hours of viral transduction, cells were split into 2g/L puromycin and selected for 72 hours before passaging for storage and downstream validation and experimentation.

STAMP editing

For stable cell STAMP fusion protein expression cells were induced with 50ng/ml (low) or 1µg/ml (high) doxycycline in DMEM for 24–72 hours, followed by Trizol extraction and Direct-zol miniprep (Zymo Research) column purification in accordance with manufacturer

protocol. Uninduced cells of the same genetic background were used as negative controls. For transient transfections, ~1 million cells were transfected with 2 μ g expression construct using Fugene HD (Promega) according to manufacturer's protocol. Upon Agilent TapeStation quantification, 500ng RNA was used as input material to make total RNA-seq libraries with either TruSeq Stranded mRNA Library Prep (Illumina) or KAPA RNA HyperPrep Kit with RiboErase (Roche) following the provided protocols. For mTOR perturbation experiments, cells were treated with 100nM Torin-1 (Cell Signaling) or DMSO vehicle control alongside 1 μ g/ml doxycycline induction and harvested for RNA after 72 hours 37C incubation.

eCLIP experiments and analysis

All STAMP fusion (RBFox2-, TIA1-, and SLBP-APOBEC1) eCLIPs were conducted following induction or transient transfections and IP was conducted using Anti-HA tag antibody - ChIP Grade (Abcam, ab91110). eCLIP experiments were performed as previously described in a detailed standard operating procedure [2], which is provided as associated documentation with each eCLIP experiment on the ENCODE portal (https://www.encodeproject.org/documents/fa2a3246-6039-46ba-b960-17fe06e7876a/@@download/attachment/CLIP_SOP_v1.0.pdf). In brief, 20 million crosslinked cells were lysed and sonicated, followed by treatment with RNase I (Thermo Fisher) to fragment RNA. Antibodies were pre-coupled to species-specific (anti-rabbit IgG) Dynabeads (Thermo Fisher), added to lysate, and incubated overnight at 4 °C. Prior to IP washes, 2% of sample was removed to serve as the paired input sample. For IP samples, high- and low-salt washes were performed, after which RNA was dephosphorylated with FastAP (Thermo Fisher) and T4 PNK (NEB) at low pH, and a 3' RNA adaptor was ligated with T4 RNA ligase (NEB). Ten per cent of IP and input samples were run on an analytical PAGE Bis-Tris protein gel, transferred to PVDF membrane, blocked in 5% dry milk in TBST, incubated with the same primary antibody used for IP (typically at 1:4,000 dilution), washed, incubated with secondary HRP-conjugated species-specific TrueBlot antibody (Rockland), and visualized with standard enhanced chemiluminescence imaging to validate successful IP. Ninety per cent of IP and input samples were run on an analytical PAGE Bis-Tris protein gel and transferred to nitrocellulose membranes, after which the region from the protein size to 75 kDa above protein size was excised from the membrane, treated with proteinase K (NEB) to release RNA, and concentrated by column purification (Zymo). Input samples were then dephosphorylated with FastAP (Thermo Fisher) and T4 PNK (NEB) at low pH, and a 3' RNA adaptor was ligated with T4 RNA ligase (NEB) to synchronize with IP samples. Reverse transcription was then performed with AffinityScript (Agilent), followed by ExoSAP-IT (Affymetrix) treatment to remove unincorporated primer. RNA was then degraded by alkaline hydrolysis, and a 3' DNA adaptor was ligated with T4 RNA ligase (NEB). qPCR was then used to determine the required amplification, followed by PCR with Q5 (NEB) and gel electrophoresis to size-select the final library. Libraries were sequenced on the HiSeq 2000, 2500, or 4000 platform (Illumina). Each ENCODE eCLIP experiment consisted of IP from two independent biosamples, along with one paired size-matched input (sampled from one of the two IP lysates before IP washes). Reproducible eCLIP peaks were called using the latest release of the core pipeline (<https://github.com/yeolab/eclip>), followed

by a peak merging sub-workflow to identify reproducible peaks (https://github.com/YeoLab/merge_peaks).

RNA-seq analysis

Bulk RNAseq was sequenced single-end 100nt and trimmed using cutadapt (v1.14.0). Trimmed reads were filtered for repeat elements using sequences obtained from RepBase (v18.05) with STAR (2.4.0i). Reads that did not map to repeats were then mapped to the hg19 assembly with STAR, sorted with samtools (v1.5) and quantified against Gencode (v19) annotations using Subread featureCounts (v1.5.3). Genes with zero counts summed across all samples were removed prior to performing differential expression analysis using DESeq2 (v1.26.0) [3].

Differential Expression (DESeq2-Supplementary table 3): To calculate differential expression from RNA-seq data, we used DESeq2 (v1.26.0), which uses a negative binomial regression model and Bayesian shrinkage estimation dispersions and fold change to estimate differentially expressed genes (Love, et al. 2014). Significance of logarithmic fold changes are determined by a Wald test to approximate p-values, and genes passing an independent filtering step are adjusted for multiple testing using the Benjamini-Hochberg procedure to yield a false discovery rate (FDR). Genes with an FDR less than 0.05 were considered statistically significant.

SAILOR calls for C-to-U edits

Resulting BAM files were each used as inputs to SAILOR (v1.1.0) to determine C>U edit sites across the hg19 assembly. Briefly described, SAILOR filters potential artifacts and known SNPs (dbSNP v147) and returns a set of candidate edit sites evidenced by the number of C>U conversions found among aligned reads. We used an adapted Bayesian “inverse probability model” [4] to identify high-confidence A-to-I editing sites from the RNA-seq data, where a confidence value based on the number of reads is associated with each predicted site. Sites were transformed into broader “window” by opening a 51-nucleotide window centered on each site.

Edit distribution, EPKM and ϵ score method details:

We describe an “ ϵ score” fraction formula: $\epsilon \text{ score} = \sum_{p=1}^i \left(\frac{\sum_{cu=0}^m Y_{cu}}{\sum_{c=1}^n Y_c} \right)$ where i represents

the number of C positions p in a given coordinate window, with Y_{cu} and Y_c representing the depth of C>U coverage m and total coverage n at each position, respectively, which considers read coverage, edit frequency (ie. how often a C>U conversion is found) and edit potential (ie. how C-rich a given region is). To find the ϵ score for a given window, we calculated the ratio between the number of (post-SAILOR-filtered) C>U read conversions to the total (post-SAILOR-filtered) coverage across every C found within the window.

To calculate Edits per kilobase of transcript per million mapped reads (EPKM) per gene, we used cumulative edit counts (T coverage over each edit site called) as determined by SAILOR (v1.1.0). We summed region-specific (either CDS or CDS+3'UTR as defined by hg19 v19 Gencode annotations) edit counts per gene and divided by “per million” mapped

read counts to either CDS or CDS+3UTR, respectively for all genes with read counts greater than 0 as defined by Subread featureCounts (v1.5.3). We then normalized this number to the length of either the CDS or CDS+3'UTR of each gene in kilobases (kb). To assess the relationship between RPS2-STAMP and mRNA translation, we compared these per gene EPKM values to normalized read units (Reads per kilobase of transcript per million mapped reads, RPKM) for ribosome protected transcripts assessed in Ribo-seq (GSE94460) and polysome-seq (GSE109423). For these analyses we included all genes that with detected read counts in either our RPS2-STAMP or ribosome-occupancy datasets.

Edit-cluster identification and de-noising

De-noising of STAMP edit data was implemented via a combination of filters designed to retain high-confidence STAMP-edited regions, followed by merging the resulting sites into coherent “peaks.” The first filter (Poisson-based filter) models the number of edited Cs relative to total C coverage as a Poisson process. Given that total edit count on any given gene correlates with expression of that gene, a gene-specific background proportion of edited Cs due to off-target effects is also assumed. By dividing the total number of C>T conversions by the total number of reads at C positions for each gene, a Poisson parameter is established for each gene, representing this background proportion. Then each edit site is individually evaluated by whether its proportion of edited Cs falls enough far to the right on its own gene’s Poisson distribution, using a baseline p-value of 0.05 with a Bonferroni correction based on the number of edit sites being evaluated on that gene, with increased stringency achieved by further dividing this per-gene adjusted p-value by a constant factor. The second approach (score-based filter) makes use of the per-site beta-distribution-derived confidence score described earlier, filtering out any edit sites with a score less than 0.999. The final approach (isolated site filter) is based on the observation that STAMP sites overlapping with the most confident eCLIP peaks tend to be found in clusters rather than isolated, and as such any edit sites with zero neighboring sites within 100 bp in either direction are filtered out. STAMP edit-clusters are generated by merging sites found within 100 bp of each other using bedtools. We performed de-novo motif finding using HOMER (v4.9.1).

Peaks exhibiting $I2fc > 2$ and $I10p > 3$ from C-Terminus RBFOX2-APOBEC fusion eCLIP data were shuffled within the 5'UTR, CDS, and 3'UTR regions of their respective genes, over 40 permutations. These peak permutations were then expanded by 200 bp on each flank and intersected with de-noised STAMP edit-clusters. The same flank-expansion and intersection was conducted for the original experimentally-derived eCLIP peaks. Six different versions of STAMP site “de-noising” are reflected by the x axis labels, where the decimal value reflected the confidence score used for filtering, and the “filtered” suffix reflects application of the additional isolated site filter.

Figure 2C–E: RPS2-STAMP at 0, 50, and 1000ng doxycycline treatments compared to corresponding control-STAMP datasets were compared across genesets taken from GSE94460 and [5]. Similar comparisons were done using normalized occupancy ratios from [6] (using X3 values, which closely approximate native ribosome occupancy levels in 293T cells). Figure 2F: Metagene profiles comparing edits (conf = 0.5) in RPS2-, RBFOX2-,

and control-STAMP were generated using metaPlotR (<https://github.com/olarerin/metaPlotR>) with the highest occupancy via ribo-seq transcripts from GSE112353, with the top quartile of expressed transcripts being used, although no expression filtering or transcript-to-gene mapping was needed as transcript-level annotations were required (Q1 n=4,677). Figure 2G: Metagene plots were generated in similar fashion to Figure 3G, comparing all replicates of Torin-1 treated RPS2-STAMP, RPS2-STAMP vehicle treated. Figure 2H: From GSE94460, genes were ranked in descending order according to their replicate-averaged TPM-normalized occupancy counts. To consolidate annotations, transcripts that were found with the highest occupancy were kept. Additionally, only genes included in both our analysis (minimally expressed protein coding genes TPM>0 in either RPS2- or control-STAMP, n=16,128) and the GSE112353 dataset (n=19,724) were used. The remaining genes (n=15,485) were quartiled according to occupancy score, such that “quartile 1” represents genes with the highest ribosome occupancy. EPKM values across CDS and 3’UTR exons within these quartiles were compared using a Wilcoxon rank-sum test to determine significance. Figure S2G: Similar to Figure 2H, using Torin-1 treated control-STAMP, vehicle treated control-STAMP.

Irreproducible Discovery Rate (IDR)

Irreproducible Discovery Rate (IDR) was employed to determine reproducible edit windows between experimental replicates (Li Q et al., *Annals of Applied Statistics*. 2011). After pre-filtering SAILOR outputs for a minimum confidence score (> 0.5), we created 51nt windows around candidate C>U sites and calculated reproducibility scores for each window using IDR (v2.0.2). Scaled score ($-125 * \log_2(\text{IDR_score})$) were converted to linear values and plotted, with unscaled scores > 0.05 considered as reproducible sites.

RNA Isolation and PolyA selection for Nanopore and PacBio Sequencing

At 80% confluency in 10cm plates, cells were washed with PBS and harvested in 1mL of TRIzol reagent (Thermo Fisher) or Direct-zol kit with DNase treatment (Zymo Research). Total RNA was extracted following the manufacturer’s protocol. 20ug of total RNA was poly-A selected using a poly-A magnetic resin kit (NEB E7490L). RNA was then analyzed by high-sensitivity RNA TapeStation (Agilent #5067–5579) to confirm poly-A selection and RNA quality.

Direct cDNA Nanopore Sequencing

100ng of poly-A selected RNA was used as input for the Nanopore direct cDNA sequencing kit (SQK-DCS109). cDNA was prepared following the manufacturer’s protocol. Sequencing was carried out on using Oxford Nanopore PromethION flow cells (FLO-PRO002) for ~48 hours. Data was base called in real time on the PromethION Guppy base callers with the high accuracy setting. Total reads (in millions) were: RBFOX2=24.9, APOBEC_control= 8.4.

Nanopore Read Base and Edit Calling

All Nanopore reads were aligned to both hg19 and ENSEMBL’s cDNA reference genomes using Minimap2 [7] with default RNA parameters. These alignments are referred to

genomic and cDNA respectively. Edits were called using Bcftools mpileup with settings “-Q 5 -d 8000 -q 1” followed by filtering each position for reference C positions on the appropriate strand. cDNA alignments were assumed to be positive stranded and genome alignments were intersected with gene annotations to determine strand. Sites with ambiguous strand information and/or fewer than 10 reads were removed. Edit fractions were determined for sites with C to U mutations by the fraction (# of mismatches)/(# of mismatches + # of matches). Confidence scores and SNP removal were done via custom implementation of the SAILOR scripts. A final list of RBFOX2 -STAMP sites was made by subtracting all sites found in the control-STAMP with a confidence score of 0.99 or greater. Isoform specific binding were detected by summing the number of RBFOX2 unique sites and all sites identified in the control-SAMP The top two expressing isoforms, as determined by average coverage across C positions with at least 10 reads, were selected for further analysis and isoforms comparing the largest difference in edits were compared by hand.

Direct cDNA PacBio sequencing

Technical triplicate samples containing 1 µg total RNA were extracted from HEK293T cells expressing control-STAMP and HEK293T, cells expressing the RBFOX2-STAMP fusions, and following 1µg/ml doxycycline induction for 72 hours. RNA extraction was completed using Direct-zol, Zymogen. All STAMP samples were assayed for quality and the all sample RNA integrity numbers (RINs) were greater than 9. Long read cDNA libraries were prepared according to the PacBio Iso-Seq Express protocol with 300 ng of total RNA and amplified for 13–15 cycles with the following forward and reverse primers:

Forward: 5'- GGCAATGAAGTCGCAGGGTTG - 3'

Reverse: 5'- AAGCAGTGGTATCAACGCAGAG – 3'

The double stranded cDNA for each sample was converted to sequencing libraries as recommended (PacBio SMRTbell Express Template Prep Kit 2.0) but with separate barcoded adapters for each sample (PacBio Barcoded Overhang Adapter Kit).

All of the samples were pooled in an equimolar fashion and sequenced on a SMRTcell 8M with the PacBio Sequel II instrument (2.0 chemistry/2.1 polymerase with 2 hour pre-extension and 30 hour movie times). After barcodes were demultiplexed, the initial data was used to re-balance the pooling by barcode counts before further sequencing. In total, the samples were sequenced over 5 SMRTcell 8M. The PacBio Sequel II system was used for all sample sequencing.

Following sequencing, the circular consensus sequence (CCS) reads for each set of technical replicates were processed using the Isoseq v3 pipeline [8] (<https://github.com/PacificBiosciences/IsoSeq>) to generate full-length non-concatamer reads in fasta format. For this step, software package lima v2.0.0 was used with parameters: --isoseq and --dump-clips. In addition, isoseq3 v3.4.0 refine was used with parameters: --require-polya. Fasta files for each set of technical replicates were the pooled together and the full-length non-concatamer reads for each sample were aligned to the hg19 reference genome using minimap2 v2.17-r941 using parameters: --ax-splice, -uf, --secondary=no, -t 30. Cupcake v18.1.0 ([*Nat Methods*. Author manuscript; available in PMC 2021 November 07.](https://</p></div><div data-bbox=)

github.com/Magdoll/cDNA_Cupcake/wiki) script `collapse_isoforms_by_sam.py` was then run using the pooled full-length non-concatemer fasta file and aligned SAM file for each sample with parameter `--dun-merge-5-shorter` to collapse redundant isoforms. This step was completed to collapse high quality isoforms into unique isoforms informed by genome alignment. Following this, SQANTI3 v1.6 [8] (<https://github.com/ConesaLab/SQANTI3>) script `sqanti3_qc.py` was used to compare the collapsed isoform results from Cupcake to the Gencode hg19 (v19) annotation in order to characterize the collapsed isoforms.

Edit/C were quantified for each sample using the SAILOR computational tool without filtering reads for RBFOX2-APOBEC1 and APOBEC1-control samples. Edited positions having a confidence score of greater than or equal to 0.99 were then used to elucidate motifs using HOMER tool `findMotifsGenome.pl` v4.9.1 [9]

A custom script was generated in order to quantify the percent edited reads in the 3' exonal region for each sample. Only previously annotated genes were considered, using the Gencode hg19 (v19) annotation as the reference. For each gene, the isoforms associated with the gene were first determined based on assignment by the SQANTI3 isoform classification pipeline. Only genes with two or more isoforms were considered. Following this, the reads associated with each isoform were determined and categorized using the `.group.txt` file generated by Cupcake. Samtools v1.9 tool `bamtobed` was used to generate a BED file based on the aligned reads for each sample. For each sample, start and end coordinates for each read associated with the gene were extracted from the BED file and used to group reads into bins based on the coordinate of the 3' end of each read, applying a leniency of a 10bp window. Only bins corresponding to the dominant 3' exon start site were considered in order to filter for bins that would support instances of alternative polyadenylation (APA). Edits in a read were counted across the region of a read between the dominant 3' exon start site and the end site corresponding to the respective bin. Edits located at potential SNP positions (positions where 50% of the reads in the bin contained an edit) were not considered. The proportion of reads containing one or more edits within the selected region corresponding to each respective bin was then quantified. Further filtering involved only comparing the two bins with the most reads for each gene and filtering out genes in which the bin with the most reads had more than five times the number of reads in the second bin.

Single cell RNA-seq

For the single cell RNA sequencing of transduced cells. Following 72 hours of doxycycline treatment (1 $\mu\text{g}/\text{mL}$), cells were trypsinized (TrypLE, Invitrogen), counted and resuspended at a density of 1,000 cell/ μL in 0.04% BSA in PBS. Single cells were processed through the Chromium Single Cell Gene Expression Solution using the Chromium Single Cell 3' Gel Bead, Chip, 3' Library and 3' Feature Barcode Library Kits v3 (10X Genomics) as per the manufacturer's protocol. Sixteen thousand total cells were added to each channel for a target recovery of 10,000 cells. The cells were then partitioned into Gel Beads in Emulsion in the Chromium instrument, where cell lysis and barcoded reverse transcription of RNA occurred, followed by amplification with the addition of "Feature cDNA Primers 1" (for the mixed RBFOX2:TIA1-STAMP), fragmentation, end-repair, A-tailing and 5' adaptor and sample index attachment as indicated in the manufacturer's protocol for 3' expression capture. 3'

feature barcode libraries were prepared as described by the manufacturer's protocol, following cDNA amplification, the Ampure cleanup supernatant was saved, amplified with Feature and Template Switch Oligo primers and finally indexed. Agilent High Sensitivity D5000 ScreenTape Assay (Agilent Technologies) was performed for QC of the libraries. 3' polyA and feature libraries were sequenced on an Illumina NovaSeq 6000. For 3' polyA demultiplexing, alignment to the hg19 and custom hg19 + lentiviral-genes transcriptomes and unique molecular identifier (UMI)-collapsing were performed using the Cellranger toolkit (version 2.0.1) provided by 10X Genomics. Cells with at least 50,000 mapped reads per cell were processed. Analysis of output digital gene expression matrices was performed using the Scanpy v1.4.4 package [10]. Matrices for all samples were concatenated when necessary and all genes that were not detected in at least 0.1% of single cells were discarded. Cells with fewer than 1,000 or more than 7,000 expressed genes as well as cells with more than 50,000 unique transcripts or 20% mitochondrial expressed genes were removed from the analysis. The only exception for these filters was for the NPC:HEK293T samples for which only cells with over 25% mitochondrial genes were filtered out and cell doublets were removed with Scrublet. Transcripts per cell were normalized to 10,000, added a unit and logarithmized ("ln(TPM+1)") and scaled to unit variance (z-scored). Top 2,000 variable genes were identified with the filter_genes_dispersion function, flavor='cell_ranger'. PCA was carried out, and the top 40 principal components were retained. With these principal components, neighborhood graphs were computed with 10 neighbors and standard parameters with the pp.neighbors function. Single cell edits were called by first computing the MD tag from Cellranger outputs (possorted_genome_bam.bam) using Samtools calmd and splitting every read according to their cell barcode. SAILOR, ϵ score, region EPKM and motif analysis were run for each cell (or the aggregate of reads for all cell barcodes within defined Louvain clusters) in similar fashion to bulk RNAseq. Reads belonging to each cluster of barcodes were combined using a custom script and treated similarly. Analysis of output digital gene edit matrices was performed using the Scanpy v1.4.4 package [10]. Matrices for all samples were concatenated and all genes that were not edited in at least 2 single cells were discarded, leaving 1,061, 1,053, 1,748, 1,542, 1,949 and 1,862 edited genes for further analyses for NPC:HEK293T-control-, NPC:HEK293T-RBFOX2-, HEK293T-control-, HEK293T-RBFOX2:TIA1-, HEK293T-RPS2-STAMP and RPS2-cluster, respectively. Cells with fewer than 10 edited genes were removed from the analysis. EPKM or ϵ scores for each cell were normalized to 10,000, added a unit and logarithmized ("ln(TPM+1)") and scaled to unit variance (z-scored). PCA was carried out, and the top 40 principal components were retained. With these principal components, neighborhood graphs were computed with 10 neighbors and standard parameters with the pp.neighbors function. Louvain clusters were computed with the tl.louvain function and standard parameters. Following visual inspection, subsets of Louvain clusters were merged guided by their overlap (or lack thereof) with control-STAMP cells in order to define RBP-specific clusters. Single cell and mean ϵ scores per sample heatmaps were generated with the pl.heatmap and pl.matrixplot functions, respectively. Differentially edited genes were determined for each set of Louvain (or modified) clusters with the tl.rank_gene_groups function (method='wilcoxon').

Figure 4G: Edits were called from groups of randomly selected RBFOX2-capture sequence barcodes (n=1..49, 100, 200, 300, 400, 500, 600, 700, 800, 844 cells) and processed using SAILOR pipeline. To discover whether or not sites are globally enriched for known binding motifs, we re-calculated the confidence score using the same e score (number of C>U read conversions over the total coverage across all C's within a window) across all 51nt windows surrounding each candidate edit site and filtered these windows using various scores (0.99 and 0.999). We performed de-novo motif finding using HOMER (v4.9.1) using these filtered windows and a shuffled background for each UTR, CDS, intron, and total genic region (findMotifs.pl foreground.fa fasta outloc -nofacts -p 4 -rna -S 20 -len 6 -noconvert -nogo -fasta background.fa), resulting in a set of fasta sequences corresponding to each 51nt edit window as well as a corresponding random background. This was repeated 10 times. HOMER was then run on each of the 580 real/random sequences to find enriched de-novo motifs. The most significant motif that most resembled the canonical RBFOX2 motif (UGCAUG) was then used as a pivot, and significance was re-calculated for this motif for each foreground/background group and trial (findMotifs.pl foreground.fa fasta output/ -nofacts -p 4 -rna -S 20 -len 6 -noconvert -nogo -known -fasta background.fa -mknown UGCAUG.motif).

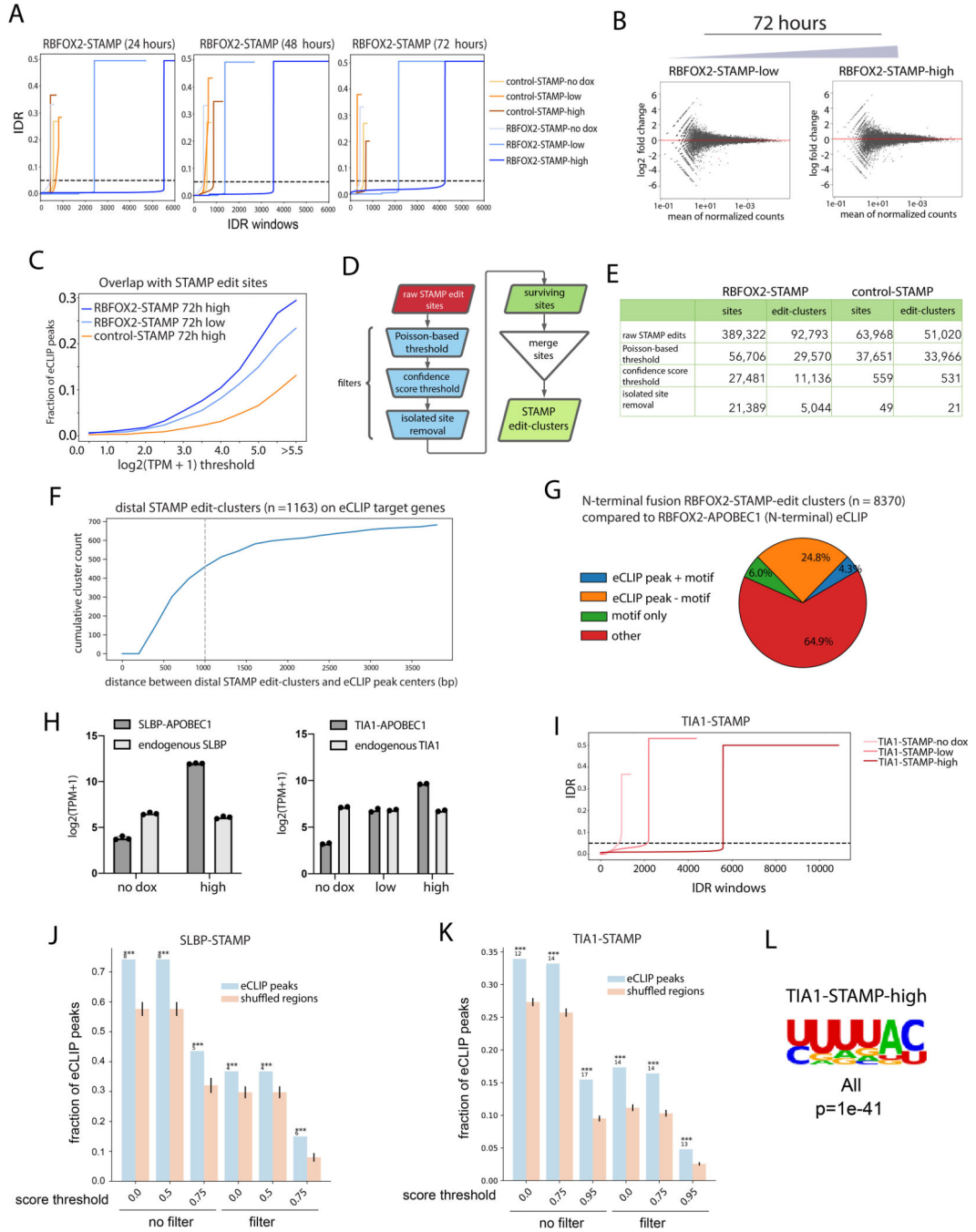
Data availability

Raw and assembled sequencing data from this study have been deposited in NCBI's Gene Expression Omnibus (GEO) under accession codes GSE155729. Processed edit coordinates are available in Supplemental Tables S1, S2, and S4. Differential edit and gene expression data are available in Supplemental Tables S3, S5–S9. Published ribosome profiling data used in this study are deposited in GEO under accession numbers GSE94460 and polysome sequencing data are deposited in GEO under accession number GSE109423.

Code availability

Source code and analysis scripts for edit quantification are available as **Supplementary Software**. Updated versions can be found at <https://github.com/YeoLab/sailor> and <https://github.com/YeoLab/STAMP>.

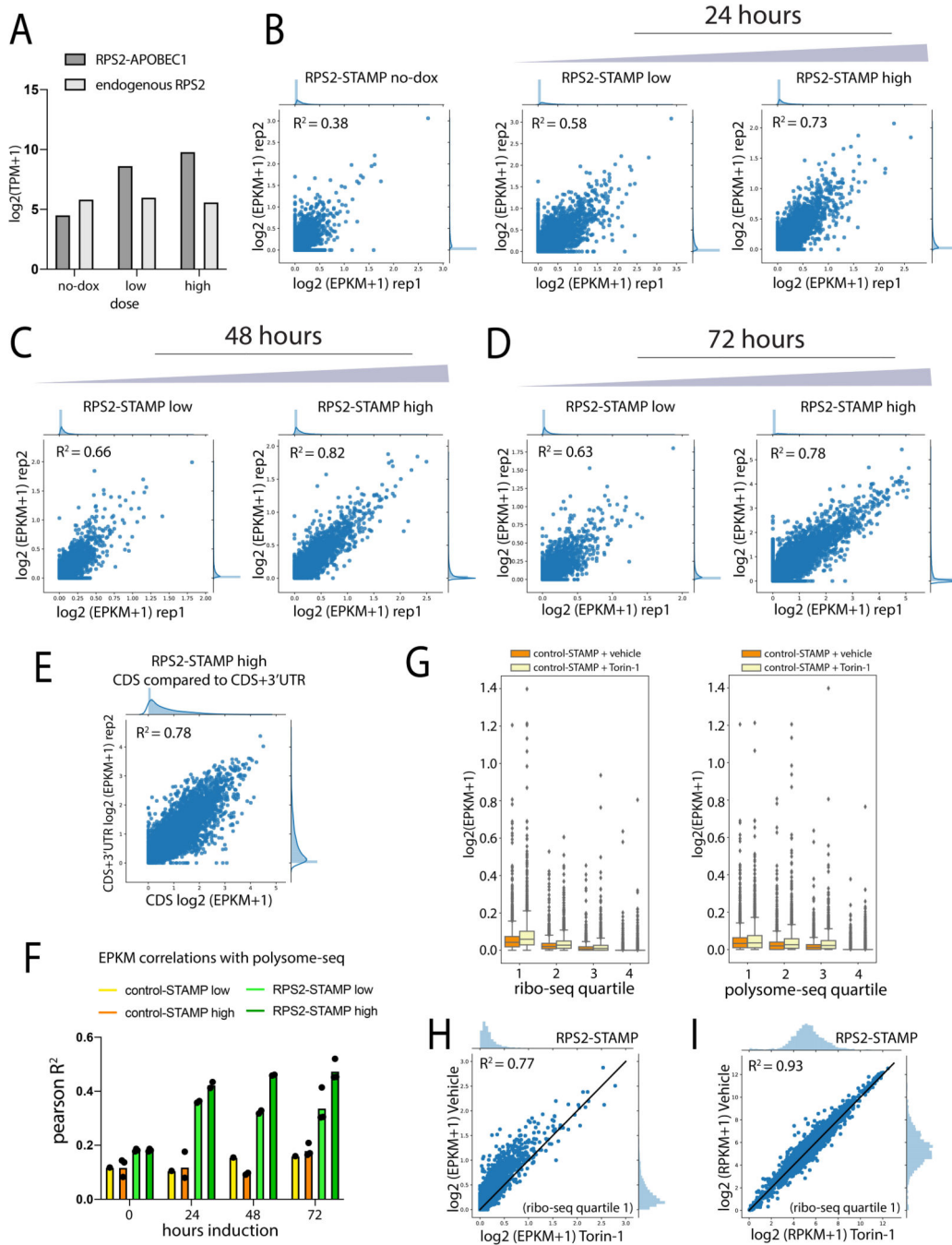
Extended Data



Extended Data Fig. 1. RBP-STAMP reproducibility and concordance with eCLIP, related to Figure 1.

A) Irreproducible Discovery Rate (IDR) analysis comparing 0.5 confidence edit windows for increasing levels of RBFOX2-STAMP at 24, 48 and 72 hours. **B)** Differential expression (DESeq2) analysis of RBFOX2-STAMP for increasing levels of RBFOX2-STAMP at 72 hours. **C)** Fraction of RBFOX2-APOBEC1 eCLIP peaks overlapping low and high induction RBFOX2-STAMP edit sites at increasing expression (TPM) thresholds. **D)** STAMP edit-site

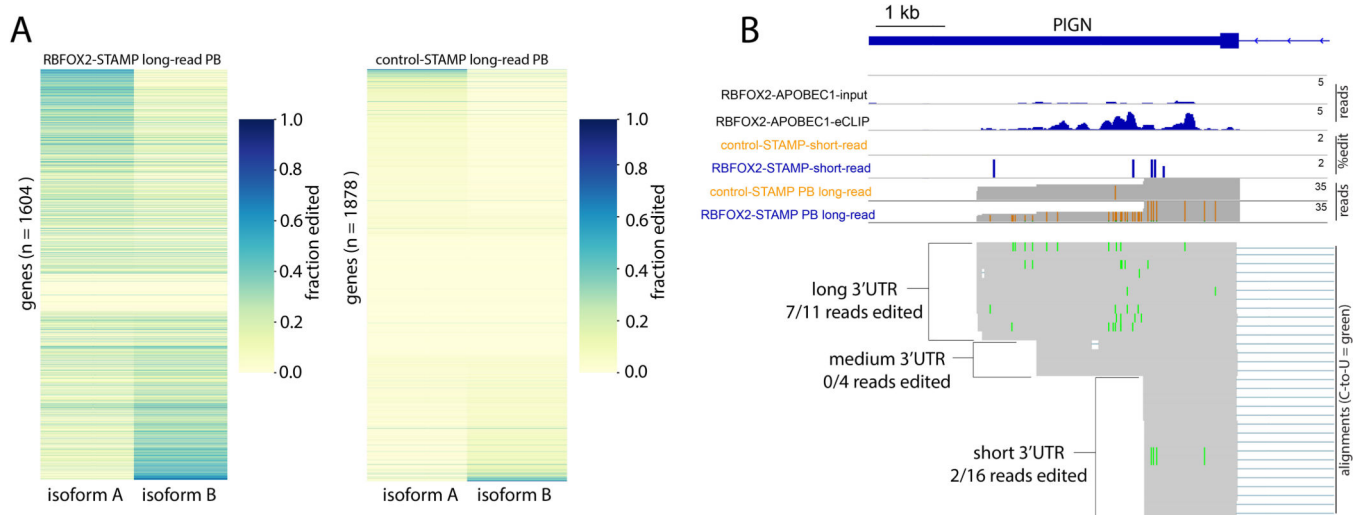
filtering and cluster-calling workflow. **E)** Number of control- and RBFOX2-STAMP edit sites and clusters retained after each filtering step in D. **F)** Cumulative distance measurement from RBFOX2-STAMP distal edit-clusters to eCLIP peaks on targets genes. **G)** Pie chart showing the proportion of N-terminally fused RBFOX2-APOBEC1 STAMP edit-clusters overlapping with either 1) RBFOX2-APOBEC1 N-terminal fusion high-confidence eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over input) containing the conserved RBFOX2 binding motif (GCAUG), 2) equally stringent eCLIP peaks not containing the conserved motif, 3) the conserved motif falling outside of eCLIP peaks, or 4) neither eCLIP peaks nor conserved motifs. **H)** Quantification of expression from no dox (0ng/ml) low (50ng/ml) or high (1 μ g/ml) doxycycline induction of SLBP-APOBEC1 and TIA1-APOBEC1 fusions compared to endogenous expression. **I)** Irreproducible Discovery Rate (IDR) analysis comparing 0.5 confidence level edit windows for increasing levels of TIA1-STAMP at 72 hours. **J)** Fraction of SLBP eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over size-matched input, reproducible by IDR) with SLBP-STAMP edit-clusters, compared to size-matched shuffled regions, calculated at different edit site confidence levels before and after site filtering (see Materials and Methods for filtering procedure). Numbers atop bars are Z-scores computed comparing observed with the distribution from random shuffles. *** denotes statistical significance at $p = 0$, one-sided exact permutation test. **K)** Fraction of TIA1-APOBEC1 eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over size-matched input) with TIA1-STAMP edit-clusters, compared to size-matched shuffled regions, calculated at different edit site confidence levels before and after site filtering (see Materials and Methods for filtering procedure). Numbers atop bars are Z-scores computed comparing observed with the distribution from random shuffles. *** denotes statistical significance at $p = 0$, one-sided exact permutation test. **L)** Motif enrichment using HOMER and shuffled background on TIA1-STAMP edit-clusters.



Extended Data Fig. 2. Ribosome-STAMP reproducibility and response to mTOR pathway perturbations, related to Figure 2.

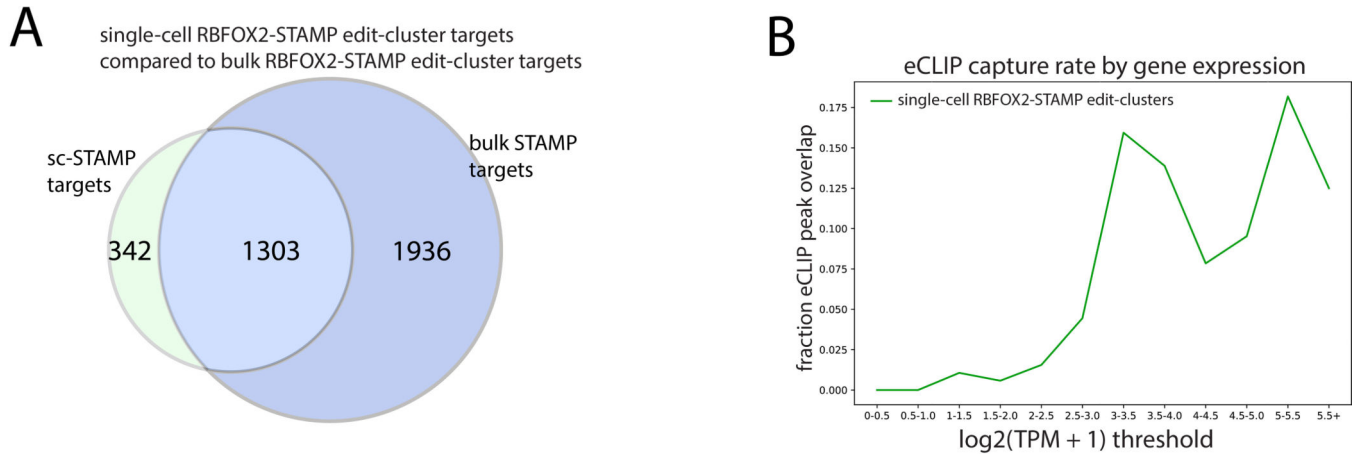
A) Quantification of expression from no dox (0ng/ml) low (50ng/ml) or high (1µg/ml) doxycycline induction of RPS2-APOBEC1 fusion compared to endogenous expression. **B-D)** Scatterplot comparisons of CDS+3'UTR EPKM values from RPS2-STAMP replicate experiments showing high, dose-dependent correlation at 24 (B), 48 (C) and 72 hours (D). **E)** Scatterplot comparison of CDS EPKM values with CDS+3'UTR EPKM values for RPS2-STAMP. **F)** Pearson R^2 values for low and high induction control- or RPS2-STAMP

EPKM compared to poly-ribosome-enriched polysome-seq RPKM. **G**) Comparison of EPKM from vehicle treated 72-hour high-induction control-STAMP compared to Torin-1 treated 72-hour high-induction control-STAMP showing no significant signal reduction for top ribosome occupied quartile genes containing Torin-1 sensitive TOP genes as detected by ribo-seq (Q1 $p = 1.0$, $n = 3589$ genes, Wilcoxon rank-sum one-sided) and polysome profiling (Q1 $p = 1.0$, $n = 3589$ genes, Wilcoxon rank-sum one-sided). **H**) Scatterplot comparison of CDS+3'UTR EPKM values on ribo-seq top quartile genes ($n = 3589$) for Torin-1 treated and vehicle treated RPS2-STAMP 72-hour high ($1\mu\text{g/ml}$) doxycycline inductions as in Figure 2H. **I**) Scatterplot comparison of CDS+3'UTR RPKM values on ribo-seq quartile-1 genes ($n = 3589$) for Torin-1 treated and vehicle treated RPS2-STAMP 72-hour high ($1\mu\text{g/ml}$) doxycycline inductions.

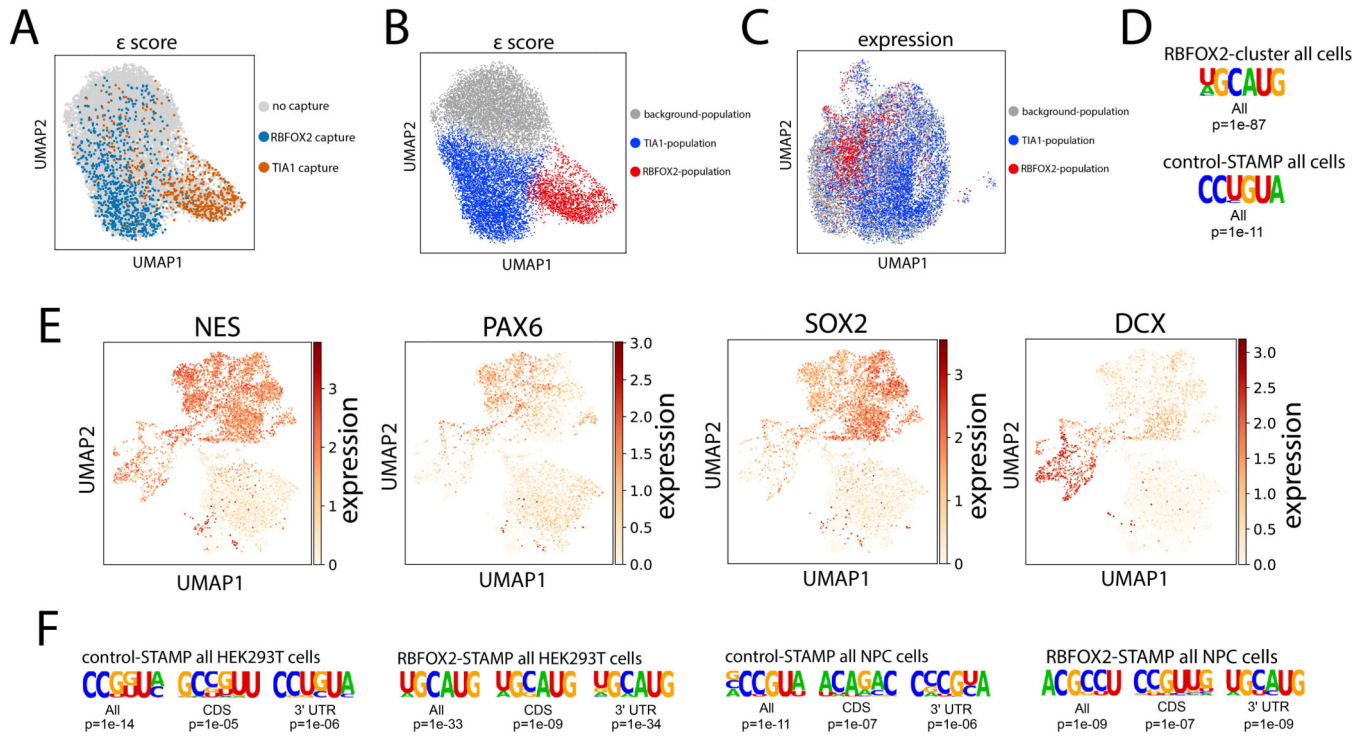


Extended Data Fig. 3. Long-read STAMP reveals isoform specific binding profiles, related to Figure 3.

A) Heatmap of control- and RBFOX2-STAMP edit fractions calculated from the final exon of all detected primary and secondary alternative polyadenylation (APA) isoforms meeting coverage criteria (see materials and methods). **B**) IGV tracks showing RBFOX2-APOBEC1 eCLIP peaks, control- and RBFOX2-STAMP short-read edit clusters, compared to control- and RBFOX2-STAMP long-read (PB) alignments on long, middle and short APA isoforms of the target gene *PIGN*, with green colored C-to-U conversions on different isoforms.



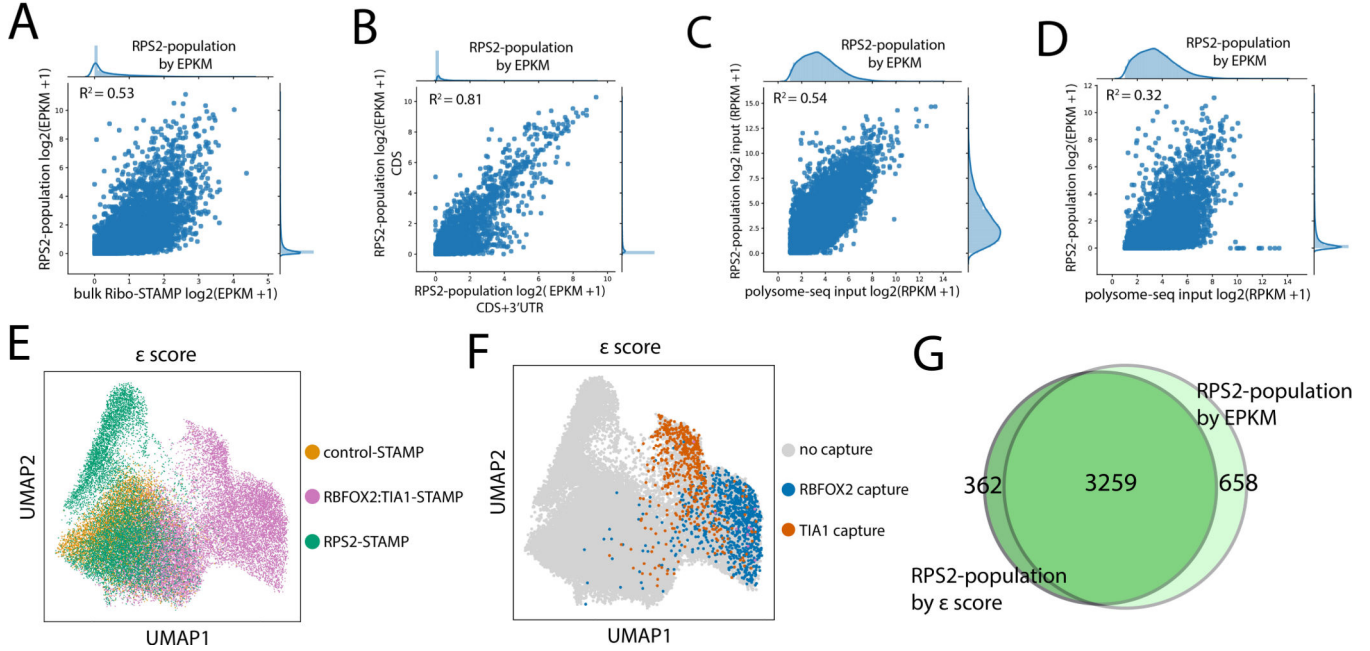
Extended Data Fig. 4. Comparison of bulk STAMP to single-cell STAMP, related to Figure 4.
A) Overlap between single-cell and bulk RBFOX2-STAMP target genes containing edit-clusters. **B)** Fraction of RBFOX2-APOBEC1 eCLIP peaks overlapping low and high induction single-cell RBFOX2-STAMP edit-clusters at increasing expression (TPM) thresholds.



Extended Data Fig. 5. Single-cell RBP-RNA interaction detection by STAMP for multiple RBPs and in multiple cell types, related to figure 5.

A) UMAP plot using ϵ score from RBFOX2-STAMP and TIA1-STAMP mixture with capture sequence RBFOX2-STAMP (blue, $n = 844$) and TIA1-STAMP cells (red, $n = 527$) highlighted. **B)** UMAP plot as in A color-coded by Louvain clustering into RBFOX2-cluster (blue), and TIA1-cluster (red), or background-cluster (gray) populations. **C)** UMAP plot of

gene expression for ϵ score Louvain clusters defined in B. **D)** Motif enrichment using HOMER from 0.99 confidence edits from combined RBFOX2-cluster and control-STAMP cells. **E)** UMAP plot showing expression of neural precursor cell markers *NES*, *PAX6*, *SOX2* and *DCX*. **F)** Motif enrichment using HOMER from 0.99 confidence edits from combined control- and RBFOX2-STAMP HEK293T and NPC cells.



Extended Data Fig. 6. Single Ribo-STAMP detects ribosome occupancy from individual cells, related to Figure 6.

A) Genome-wide comparison of CDS+3'UTR EPKM values for bulk and single-cell EPKM-derived RPS2-population. **B)** Comparison of EPKM-derived RPS2-population CDS and CDS+3'UTR EPKM values. **C)** Comparison of EPKM-derived RPS2-population total mRNA RPKM values with total mRNA RPKM values from polysome-seq input. **D)** Comparison of EPKM-derived RPS2-population CDS+3'UTR EPKM values with total mRNA RPKM values from polysome-seq input. **E)** UMAP analysis of ϵ score from merged 72-hour high-induction RPS2-STAMP (green), control-STAMP (orange) and mixed-cell RBFOX2:TIA1-STAMP (purple) single-cell experiments. **F)** UMAP plot as in E with only capture sequence RBFOX2-STAMP (blue, $n = 844$) and TIA1-STAMP cells (red, $n = 527$) highlighted. **D)** Individual cell barcode overlap for EPKM-derived and ϵ score-derived RPS2-populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Gabriella Viero at the Institute of Biophysics, CNR Unit at Trento for her helpful advice concerning analysis of ribo-seq data. We thank Dr. Todd Michael at the J. Craig Venter Institute for use of the Oxford Nanopore PromethION system. We are grateful to Dr. Kate D. Meyer for her gift of the YTH-APOBEC1 construct. We thank

Dr. Jeffrey Rothstein at Johns Hopkins University School of Medicine for the gift of the NPC cell line. We are grateful to the La Jolla Institute's Immunology Sequencing Core and the IGM Genomics Center, University of California San Diego, for use of the 10X Chromium and Illumina sequencing platforms. This work was partially supported by National Institutes of Health HG004659 and HG009889 to G.W.Y. I.A.C. is a San Diego IRACDA Fellow supported by NIH/NIGMS K12 GM068524 Award. R.J.M. was supported in part an institutional award to the UCSD Genetics Training Program from the National Institute for General Medical Sciences, T32 GM008666 and a Ruth L. Kirschstein National Research Service Award (1-F31-NS111859-01A1). K.W.B is a University of California President's Postdoctoral Fellow supported by NIH/NINDS K22NS112678 K22 Award.

References

1. Singh G, et al., The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem*, 2015. 84: p. 325–54. [PubMed: 25784054]
2. Gerstberger S, Hafner M, and Tuschl T, A census of human RNA-binding proteins. *Nat Rev Genet*, 2014. 15(12): p. 829–45. [PubMed: 25365966]
3. Van Nostrand EL, et al., Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol*, 2020. 21(1): p. 90. [PubMed: 32252787]
4. Martinez FJ, et al., Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron*, 2016. 92(4): p. 780–795. [PubMed: 27773581]
5. Ramanathan M, Porter DF, and Khavari PA, Methods to study RNA-protein interactions. *Nat Methods*, 2019. 16(3): p. 225–234. [PubMed: 30804549]
6. Wheeler EC, Van Nostrand EL, and Yeo GW, Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. *Wiley Interdiscip Rev RNA*, 2018. 9(1).
7. Van Nostrand EL, et al., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*, 2016. 13(6): p. 508–14. [PubMed: 27018577]
8. Perez-Perri JI, et al., Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat Commun*, 2018. 9(1): p. 4408. [PubMed: 30352994]
9. Calviello L. and Ohler U, Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet*, 2017. 33(10): p. 728–744. [PubMed: 28887026]
10. Ingolia NT, et al., Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 2009. 324(5924): p. 218–23. [PubMed: 19213877]
11. Lee FCY and Ule J, Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*, 2018. 69(3): p. 354–369. [PubMed: 29395060]
12. Clamer M, et al., Active Ribosome Profiling with RiboLace. *Cell Rep*, 2018. 25(4): p. 1097–1108 e5.
13. Buenrostro JD, et al., Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 2013. 10(12): p. 1213–8. [PubMed: 24097267]
14. Hwang B, Lee JH, and Bang D, Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med*, 2018. 50(8): p. 96.
15. Tang F, et al., mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 2009. 6(5): p. 377–82. [PubMed: 19349980]
16. Stoeckius M, et al., Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*, 2017. 14(9): p. 865–868. [PubMed: 28759029]
17. Shahi P, et al., Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci Rep*, 2017. 7: p. 44447.
18. Nguyen DTT, et al., HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nat Commun*, 2020. 11(1): p. 2026. [PubMed: 32332729]
19. Medina-Munoz HC, et al., Records of RNA locations in living yeast revealed through covalent marks. *Proc Natl Acad Sci U S A*, 2020.
20. Jin H, et al., TRIBE editing reveals specific mRNA targets of eIF4E-BP in *Drosophila* and in mammals. *Sci Adv*, 2020. 6(33): p. eabb8771.

21. McMahon AC, et al., TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell*, 2016. 165(3): p. 742–53. [PubMed: 27040499]
22. Lapointe CP, et al., Protein-RNA networks revealed through covalent RNA marks. *Nat Methods*, 2015. 12(12): p. 1163–70. [PubMed: 26524240]
23. Xu W, Rahman R, and Rosbash M, Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. *RNA*, 2018. 24(2): p. 173–182. [PubMed: 29127211]
24. Rahman R, et al., Identification of RNA-binding protein targets with HyperTRIBE. *Nat Protoc*, 2018. 13(8): p. 1829–1849. [PubMed: 30013039]
25. Matthews MM, et al., Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol*, 2016. 23(5): p. 426–33. [PubMed: 27065196]
26. Navaratnam N, et al., The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem*, 1993. 268(28): p. 20709–12.
27. Meyer KD, DART-seq: an antibody-free method for global m(6)A detection. *Nat Methods*, 2019. 16(12): p. 1275–1280. [PubMed: 31548708]
28. Deffit SN, et al., The *C. elegans* neural editome reveals an ADAR target mRNA required for proper chemotaxis. *Elife*, 2017. 6.
29. Washburn MC, et al., The dsRBP and inactive editor ADR-1 utilizes dsRNA binding to regulate A-to-I RNA editing across the *C. elegans* transcriptome. *Cell Rep*, 2014. 6(4): p. 599–607. [PubMed: 24508457]
30. Lovci MT, et al., Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*, 2013. 20(12): p. 1434–42. [PubMed: 24213538]
31. Yeo GW, et al., An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*, 2009. 16(2): p. 130–7. [PubMed: 19136955]
32. Ponthier JL, et al., Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. *J Biol Chem*, 2006. 281(18): p. 12468–74.
33. Van Nostrand EL, et al., CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. *Methods*, 2017. 118–119: p. 50–59.
34. Li QH, et al., Measuring Reproducibility of High-Throughput Experiments. *Annals of Applied Statistics*, 2011. 5(3): p. 1752–1779.
35. Marzluff WF, Wagner EJ, and Duronio RJ, Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet*, 2008. 9(11): p. 843–54. [PubMed: 18927579]
36. Gilks N, et al., Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol Biol Cell*, 2004. 15(12): p. 5383–98. [PubMed: 15371533]
37. Davis CA, et al., The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*, 2018. 46(D1): p. D794–D801. [PubMed: 29126249]
38. Li BB, et al., Targeted profiling of RNA translation reveals mTOR-4EBP1/2-independent translation regulation of mRNAs encoding ribosomal proteins. *Proc Natl Acad Sci U S A*, 2018. 115(40): p. E9325–E9332.
39. Yang F, et al., MALAT-1 interacts with hnRNP C in cell cycle regulation. *FEBS Lett*, 2013. 587(19): p. 3175–81. [PubMed: 23973260]
40. Zhang P, et al., Genome-wide identification and differential analysis of translational initiation. *Nat Commun*, 2017. 8(1): p. 1749. [PubMed: 29170441]
41. Tan FE, et al., A Transcriptome-wide Translational Program Defined by LIN28B Expression Level. *Mol Cell*, 2019. 73(2): p. 304–313 e3.
42. Wagner S, et al., Selective Translation Complex Profiling Reveals Staged Initiation and Co-translational Assembly of Initiation Factor Complexes. *Mol Cell*, 2020. 79(4): p. 546–560 e7.
43. Archer SK, et al., Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*, 2016. 535(7613): p. 570–4. [PubMed: 27437580]
44. Miettinen TP and Bjorklund M, Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res*, 2015. 43(2): p. 1019–34. [PubMed: 25550424]

45. Thoreen CC, et al., An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J Biol Chem*, 2009. 284(12): p. 8023–32. [PubMed: 19150980]
46. Thoreen CC, et al., A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, 2012. 485(7396): p. 109–13. [PubMed: 22552098]
47. Jain M, et al., The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*, 2016. 17(1): p. 239. [PubMed: 27887629]
48. Branton D, et al., The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 2008. 26(10): p. 1146–53. [PubMed: 18846088]
49. Ardui S, et al., Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res*, 2018. 46(5): p. 2159–2168. [PubMed: 29401301]
50. Rhoads A. and Au KF, PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*, 2015. 13(5): p. 278–89. [PubMed: 26542840]
51. Fu S, Wang A, and Au KF, A comparative evaluation of hybrid error correction methods for error-prone long reads. *Genome Biol*, 2019. 20(1): p. 26. [PubMed: 30717772]
52. Li Y, et al., A comprehensive library of familial human amyotrophic lateral sclerosis induced pluripotent stem cells. *PLoS One*, 2015. 10(3): p. e0118266.
53. Song Y, et al., irCLASH reveals RNA substrates recognized by human ADARs. *Nat Struct Mol Biol*, 2020. 27(4): p. 351–362. [PubMed: 32203492]
54. Beaudoin JD, et al., Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat Struct Mol Biol*, 2018. 25(8): p. 677–686. [PubMed: 30061596]
55. Lorenz DA, et al., Direct RNA sequencing enables m(6)A detection in endogenous transcript isoforms at base-specific resolution. *RNA*, 2020. 26(1): p. 19–28. [PubMed: 31624092]

References

1. Li Y, et al., A comprehensive library of familial human amyotrophic lateral sclerosis induced pluripotent stem cells. *PLoS One*, 2015. 10(3): p. e0118266.
2. Van Nostrand EL, et al., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*, 2016. 13(6): p. 508–14. [PubMed: 27018577]
3. Love MI, Huber W, and Anders S, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 2014. 15(12): p. 550. [PubMed: 25516281]
4. Li H, Ruan J, and Durbin R, Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*, 2008. 18(11): p. 1851–8. [PubMed: 18714091]
5. Zhang P, et al., Genome-wide identification and differential analysis of translational initiation. *Nat Commun*, 2017. 8(1): p. 1749. [PubMed: 29170441]
6. Tan FE, et al., A Transcriptome-wide Translational Program Defined by LIN28B Expression Level. *Mol Cell*, 2019. 73(2): p. 304–313 e3.
7. Li H, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 2018. 34(18): p. 3094–3100. [PubMed: 29750242]
8. Gordon SP, et al., Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. *PLoS One*, 2015. 10(7): p. e0132628. [PubMed: 26177194]
9. Heinz S, et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 2010. 38(4): p. 576–89. [PubMed: 20513432]
10. Wolf FA, Angerer P, and Theis FJ, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*, 2018. 19(1): p. 15. [PubMed: 29409532]

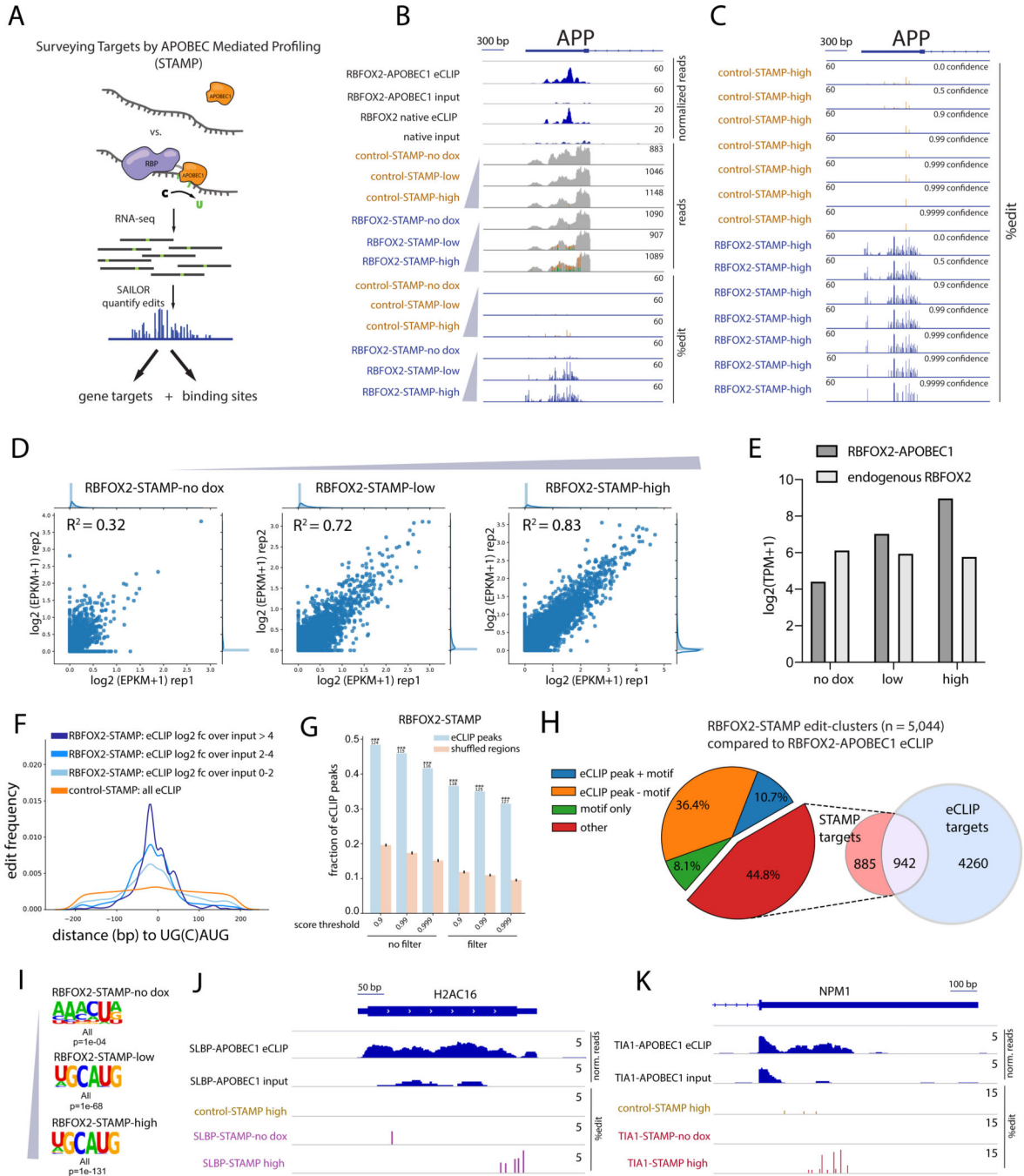


Figure 1: RBP-STAMP edits mark specific RBP binding sites.

A) Surveying Targets by APOBEC Mediated Profiling (STAMP) strategy fuses rat APOBEC1 module to an RBP of interest to deposit edits at or near RBP binding sites. C-to-U mutations from either APOBEC1-only control (control-STAMP) or RBP fusion (RBP-STAMP) can be detected by standard RNA-sequencing and quantified using our SAILOR analysis pipeline. **B)** Integrative genome viewer (IGV) browser tracks showing RBFOX2 and RBFOX2-APOBEC1 eCLIP peaks on the target gene *APP*, compared with control- and RBFOX2-STAMP signal and SAILOR quantified edit fraction for increasing induction

levels of fusions (doxycycline: 0ng = none, 50ng = low, or 1µg/ml = high, 72 hours). **C)** IGV tracks showing 72-hour high-induction control- and RBFOX2-STAMP signal on the *APP* target gene at increasing confidence levels. **D)** RBFOX2-STAMP replicate correlations for the edited read counts per target normalized for length and coverage (EPKM). **E)** Quantification of expression from no dox (0ng/ml) low (50ng/ml) or high (1µg/ml) doxycycline induction of RBFOX2-APOBEC1 fusion compared to endogenous RBFOX2 expression. **F)** RBFOX2-STAMP and control-STAMP (background) edit frequency distribution within a 400 bp window flanking RBFOX2 eCLIP binding-site motifs, split into increasing levels of log₂ fold enrichment of eCLIP peak read-density over size-matched input. **G)** Fraction of RBFOX2-APOBEC1 eCLIP peaks (log₂fc>2 and -log₁₀p>3 over size-matched input) with RBFOX2-STAMP edit-clusters, compared to size-matched shuffled regions, calculated at different edit site confidence levels before and after site filtering (see Materials and Methods for filtering procedure). Numbers atop bars are Z-scores computed comparing observed with the distribution from random shuffles. *** denotes statistical significance at p = 0, one-sided exact permutation test. **H)** Pie chart showing the proportion of filtered RBFOX2-STAMP edit-clusters overlapping either 1) RBFOX2-APOBEC1 fusion high-confidence eCLIP peaks (log₂fc>2 and -log₁₀p>3) containing the conserved RBFOX2 binding motif, 2) equally stringent eCLIP peaks not containing the conserved motif, 3) the conserved motif falling outside of eCLIP peaks, or 4) neither eCLIP peaks or conserved motifs. **I)** Motif enrichment using HOMER and shuffled background on RBFOX2-STAMP edit-clusters for increasing RBFOX2-STAMP induction levels. **J)** IGV tracks showing control- and SLBP-STAMP edit fractions at no- and high-induction (doxycycline: 0ng = none or 1µg/ml = high, 72 hours) on the target histone gene *H2AC16* compared to SLBP-APOBEC1 eCLIP. **K)** IGV tracks showing control- and TIA1-STAMP edit fractions at no- and high-induction (doxycycline: 0ng = none or 1µg/ml = high, 72 hours) on the target gene *NPM1* compared to TIA1-APOBEC1 eCLIP.

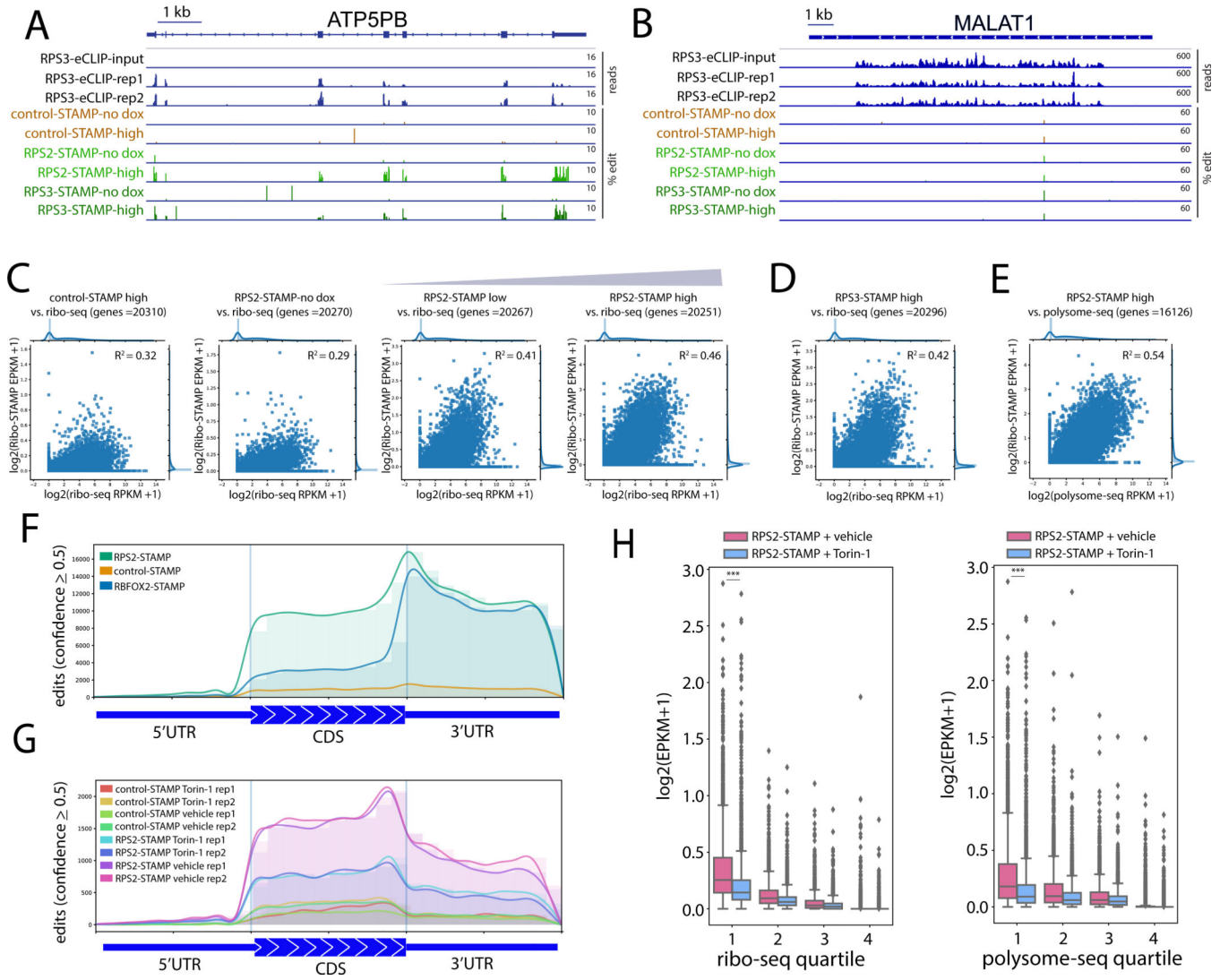


Figure 2: Ribo-STAMP edits mark highly translated coding sequences.

A) IGV browser tracks displaying coding sequence edit frequency from control, RPS2-STAMP, and RPS3-STAMP at no-induction or 72-hour high-induction on the *ATP5PB* gene locus. RPS3 eCLIP and input reads are shown for comparison. **B)** IGV browser tracks as in A on the noncoding RNA *MALAT1*, showing no enrichment for RPS3 eCLIP reads, RPS2- or RPS3-STAMP edits. **C)** Genome-wide scatterplot comparison of control- and RPS2-STAMP EPKM and ribo-seq ribosome protected fragment (RPF) RPKM for increasing levels of RPS2-STAMP. **D)** Comparison as in C with ribo-seq RPF RPKM and EPKM from RPS3-STAMP. **E)** Comparison as in C with polysome-seq RPKM and EPKM from RPS2-STAMP. **F)** Metagene plot showing edit (≥ 0.5 confidence score) distribution for high-induction RPS2-STAMP compared to control-STAMP and RBFOX2-STAMP across 5'UTR, CDS and 3'UTR gene regions for the top quartile ($n=4,931$) of ribosome occupied genes (ribo-seq). **G)** Metagene plot as in F showing edit (≥ 0.5 confidence level) distribution for vehicle-treated 72-hour high-induction RPS2-STAMP compared to replicate Torin-1 treated 72-hour high-induction RPS2-STAMP across 5'UTR, CDS and 3'UTR gene regions for the

top quartile of ribosome occupied genes. **H)** Comparison of EPKM from combined replicates (n = 2) vehicle treated 72-hour high-induction RPS2-STAMP compared to Torin-1 treated 72-hour high-induction RPS2-STAMP showing significant signal reduction for top ribosome occupied quartile genes containing Torin-1 sensitive TOP genes as detected by ribo-seq (Q1 p = 1.9 e-147, n = 3589 genes, Wilcoxon rank-sum one-sided) and polysome profiling (Q1 p = 7.7 e-108, n = 3589 genes, Wilcoxon rank-sum one-sided).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

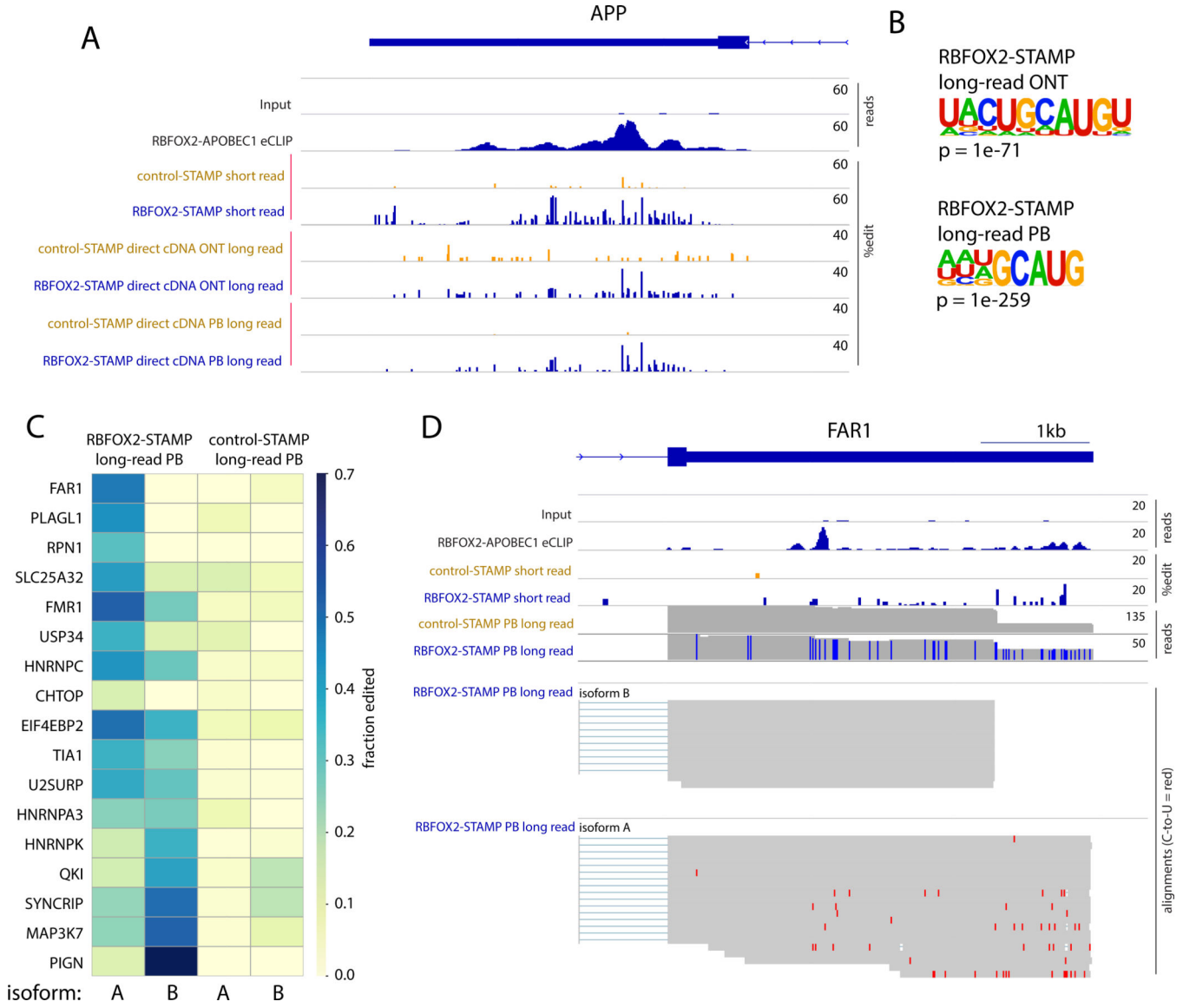


Figure 3: Long-read STAMP reveals isoform specific binding profiles.
A) IGV tracks showing RBFOX2 eCLIP peak on the target gene *APP*, compared with 72-hour high-induction control- and RBFOX2-STAMP SAILOR quantified edit fractions for both long-read (Oxford Nanopore Technologies (ONT) or PacBio (PB)) direct cDNA, and short read (NGS) outputs. **B)** Homer motif analysis of RBFOX2-STAMP long-reads (ONT and PB) for edits above 0.99 confidence. **C)** Heatmap of control- and RBFOX2-STAMP edit fractions on the 2 primary alternative polyadenylation (APA) isoforms for the top differentially edited RBFOX2-STAMP APA targets. **D)** IGV tracks showing RBFOX2-APOBEC1 eCLIP peaks, control- and RBFOX2-STAMP short-read edit frequencies, and control- and RBFOX2-STAMP long-read (PB) alignments on the 2 primary isoforms of the target gene *FAR1*, with red colored C-to-U conversions on different isoforms.

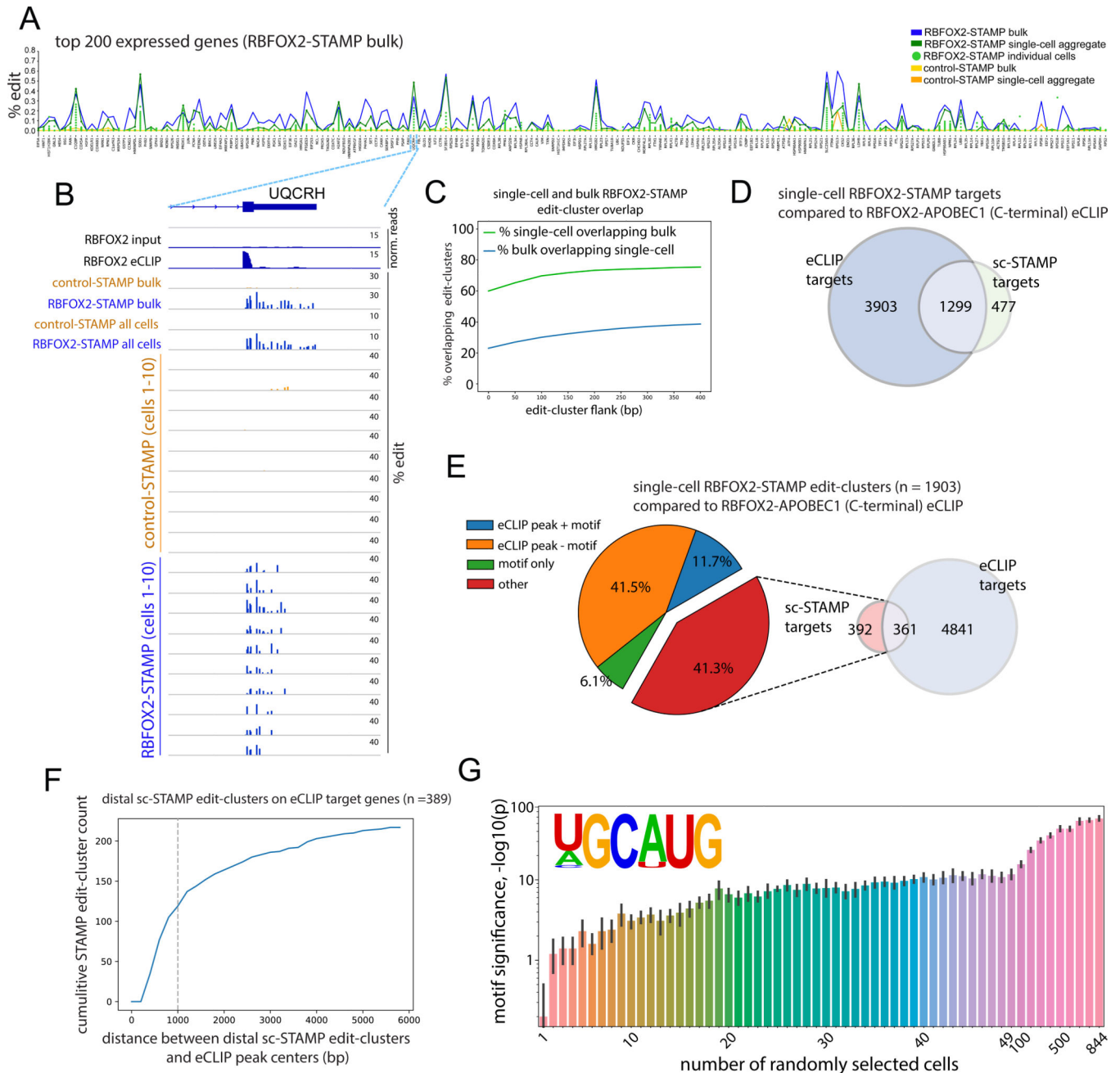


Figure 4: STAMP allows RBP binding site detection at single-cell resolution.

A) Edit fraction comparison of bulk 72-hour high-induction control- and RBFox2-STAMP with single-cell control- and RBFox2-STAMP across the top 200 genes ranked by transcripts per million (TPM) from bulk RBFox2-STAMP RNA-seq. **B)** IGV tracks showing the RBFox2 eCLIP peak on the target gene *UQCRH*, compared with RBFox2-STAMP edit fractions for the top 10 control- and RBFox2-STAMP cells ranked by summed *e* scores. **C)** Evaluation of percentage overlap between bulk and single-cell edit-clusters showing that 60–75% of single-cell edit clusters overlap bulk edit clusters over increasing cluster-flanking regions. **D)** Overlap between RBFox2-APOBEC1 eCLIP target transcripts

(peaks $\log_2fc > 2$ and $-\log_{10}p > 3$ over input) and single-cell RBFOX2-STAMP edit-cluster containing target transcripts. **E**) Pie chart showing the proportion of single-cell RBFOX2-STAMP edit-clusters overlapping either 1) RBFOX2-APOBEC1 fusion high-confidence eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over input) containing the conserved RBFOX2 binding motif (GCAUG), 2) equally stringent eCLIP peaks not containing the conserved motif, 3) the conserved motif falling outside of eCLIP peaks, or 4) neither eCLIP peaks nor conserved motifs. **F**) Cumulative distance measurement from single-cell RBFOX2-STAMP distal edit-clusters to eCLIP peaks on targets genes. **G**) $-\log_{10}$ of p-values ($n = 10$ trials) for motifs extracted by HOMER (v4.9.1) using RBFOX2-STAMP 0.99 confidence level edits from randomly sampled cells showing RBFOX2 motif detection to 1 cell resolution.

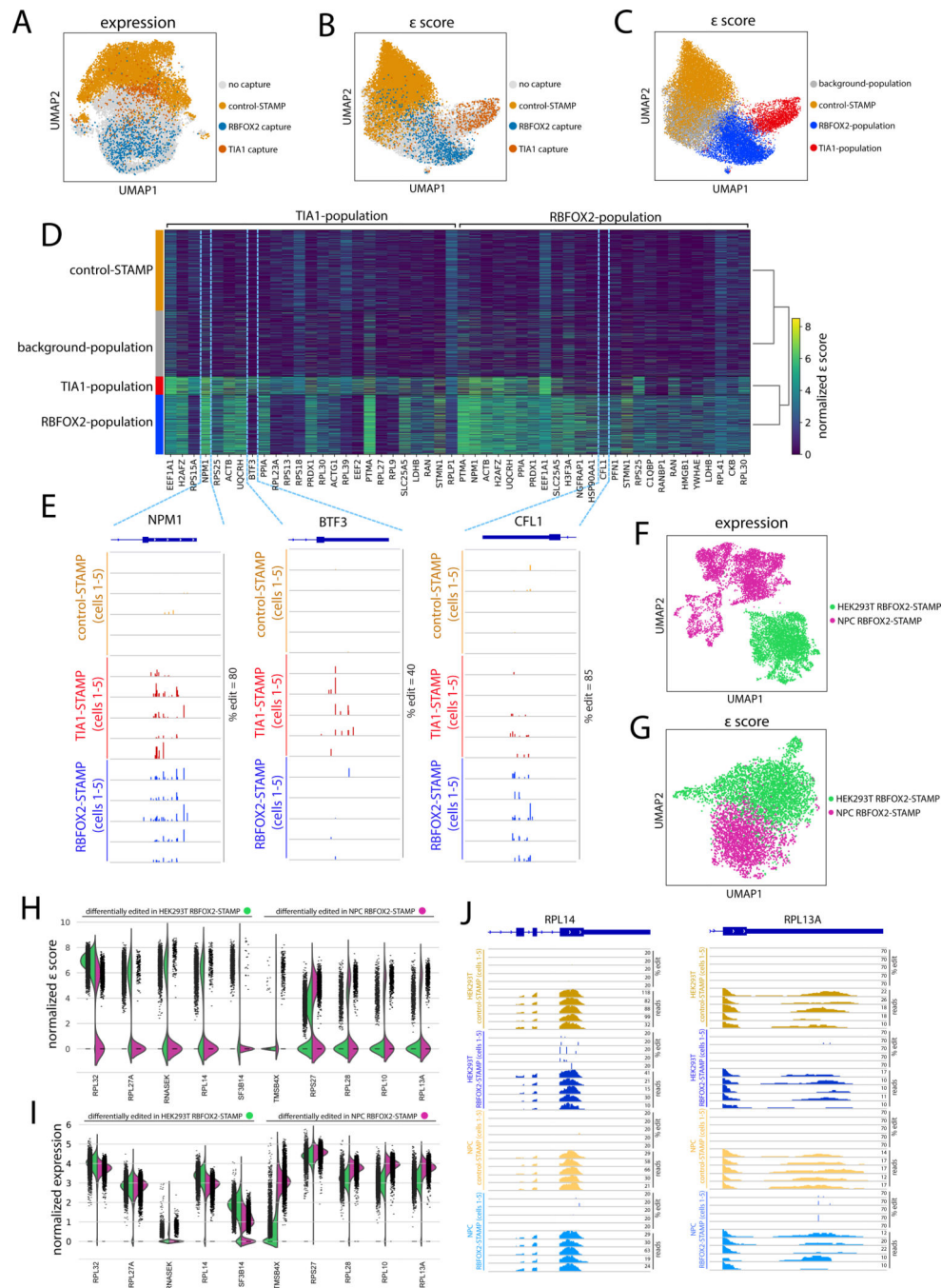


Figure 5: Deconvolution of multiple RBPs and cell-type specific targets.

A) Uniform Manifold Approximation and Projection (UMAP) analysis of gene expression from merged 72-hour high-induction control- and RBFOX2:TIA1-STAMP cells with capture sequence RBFOX2-STAMP (blue, n = 844) and TIA1-STAMP cells (red, n = 527) highlighted. B) UMAP analysis using ϵ score rather than gene expression after merging 72-hour high-induction control-STAMP cells (orange). C) UMAP plot as in B color-coded by ϵ score Louvain clustering into RBFOX2-population (blue), TIA1-population (red) and background-population (gray) populations with control-STAMP cells (orange) overlaid. D)

Heatmap of normalized ϵ score signatures for RBFOX2- and TIA1-population cells compared to control-STAMP and background cells on the top 25 differentially edited gene targets. **E)** IGV browser tracks showing SAILOR quantified edit fractions for the top 5 control-, RBFOX2-, and TIA1-STAMP cells (ranked by summed ϵ scores) on the *NPM1*, *BTF3* and *CFL1* gene targets. **F)** UMAP analysis of merged 72-hour high-induction RBFOX2-STAMP mixed NPC and HEK293T cells clustered by expression. **G)** UMAP analysis as in F using ϵ score. **H)** ϵ score distribution summarized by violin plot for HEK293T and NPC defined cell populations for the top differentially edited genes. **I)** Violin plots as in H summarizing expression rather than ϵ score. **J)** IGV browser tracks showing edit fractions and read coverage for the top 5 control- and RBFOX2-STAMP cells (ranked by summed ϵ scores) on the *RPL14* and *RPL13A* gene targets.

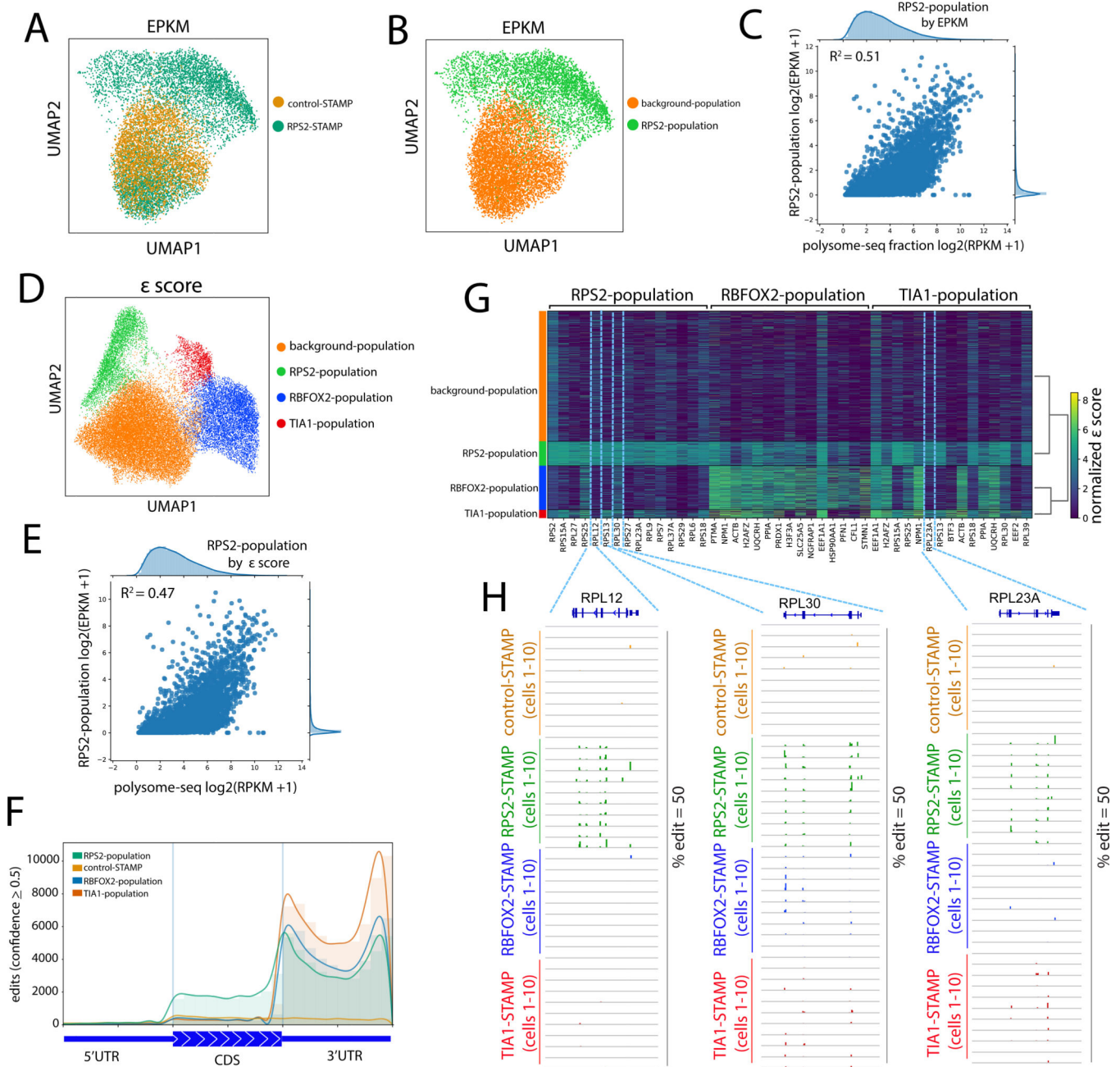


Figure 6: Ribo-STAMP reveals ribosome occupancy from individual cells.

A) UMAP analysis of EPKM for 72-hour high-induction RPS2-STAMP (green), control-STAMP (orange). **B)** UMAP analysis of cells shown in A with EPKM Louvain clustering into background-population and RPS2-population. **C)** Comparison of EPKM-derived RPS2-population CDS+3'UTR EPKM values with poly-ribosome-fraction-enriched polysome-seq RPKM values. **D)** UMAP plot color-coded by ϵ score Louvain clustering into background-cluster (orange), RBFOX2-cluster (blue), TIA1-cluster (red), and 677 RPS2-cluster (green) from merged 72-hour high-induction STAMP experiments. **E)** Comparison of ϵ score-derived RPS2-population CDS+3'UTR EPKM values with poly-ribosome-fraction-enriched

polysome-seq RPKM values. **F)** Metagene plot showing distribution for aggregate cell edits (0.5 confidence level) from control-STAMP, RPS2-cluster, TIA1-cluster, and RBFOX2-cluster cells across 5'UTR, CDS and 3'UTR gene regions for the top quartile of ribosome occupied genes. **G)** Heatmap of normalized ϵ score signatures for RPS2-population, RBFOX2-population, and TIA1-population cells compared to background cells on the top 15 differentially edited gene targets. **H)** IGV browser tracks showing edit fractions for the top 10 control-, RPS2-, RBFOX2-, and TIA1-STAMP cells (ranked by summed ϵ scores) on the *RPL12*, *RPL30* and *RPL23A* gene targets.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript