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

Kristopher Brannan
University of California San Diego
Isaac Chaim
University of California San Diego
Brian Yee
University of California San Diego
Ryan Marina
University of California San Diego
Daniel Lorenz
University of California San Diego
Kevin Dong
University of California San Diego
Assael Madrigal
University of California San Diego
Gene Yeo (geneyeo@ucsd.edu)
University of California San Diego https://orcid.org/0000-0002-0799-6037

Article

Keywords: RNA binding proteins (RBPs), RBP regulation, STAMP (Surveying Targets by APOBEC Mediated Profiling), RNA binding proteins and ribosomes

Posted Date: October 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-87224/v1

Version of Record: A version of this preprint was published at Nature Methods on May 1st, 2021. See the published version at https://doi.org/10.1038/s41592-021-01128-0.

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

Kristopher Brannan[1,2,3], Isaac A. Chaim[1,2,3], Brian A. Yee [1,2,3], Ryan J. Marina[1,2,3], Daniel A. Lorenz[1,2,3], Kevin D. Dong[1,2,3], Assael A. Madrigal[1,2,3], Gene W. Yeo[1,2,3*]

[1]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California, USA
[2]Stem Cell Program, University of California San Diego, La Jolla, California, USA
[3]Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA

*Corresponding author email: **geneyeo@ucsd.edu**

Highlights

- STAMP identifies binding sites of full-length RBPs by C-to-U RNA editing.
- STAMP coupled to long read sequencing reveals isoform specific RBP targets.
- STAMP allows cell-type specific and multiplexed-RBP target identification in single cells.
- STAMP with ribosome subunits allows detection of ribosome association at single-cell level.

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 also yields RBP target sites for full-length isoforms. Finally, conducting STAMP using small ribosomal subunits (Ribo-STAMP) allows analysis of 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]. The eukaryotic ribosome is itself composed of a collection of RBPs that can interact directly with mRNA coding sequences [3, 6]. Ribosome profiling methods such as ribo-seq, like CLIP-seq, have become a mainstay in the evaluation of transcriptome-scale translational efficiency [7, 8]. Unfortunately, CLIP and ribosome profiling experimental protocols are labor-intensive and usually require sizable amounts of input material [9, 10], and are not amenable for experiments that necessitate low input material or single cell resolution, and also are unable to resolve target transcript isoforms due to transcript fragmentation. In contrast, there has been rapid progress in single-cell measurements of chromatin accessibility [11], gene expression [12, 13], and even surface protein-levels [14, 15]. However, there is currently no available technology or documented demonstration that we can measure RBP and ribosome interaction sites on mRNA at single cell level resolution.

As a step towards lowering input requirements for RBP-recognition, recent studies have circumvented the need for immunoprecipitation by utilizing fusions of RNA-editing or RNA-modifying modules to RBPs of interest to label RNA targets. This field of antibody-free RBP-RNA interactome detection has grown steadily in recent years. The first demonstration of this general strategy, termed “RNA-tagging”, utilized fusion to the *Caenorhabditis elegans* poly(U) polymerase PUP-2 to RBPs to covalently mark target 3’ ends with poly(U) chains, which allowed transcriptome-wide target identification in *Saccharomyces cerevisiae* [16]. This poly(U) tagging was recently extended to probe RNA localization by tethering fusions to the endoplasmic reticulum or to the mitochondria [17]. These poly(U) tagging approaches unfortunately cannot provide information about binding site locations within identified target RNAs and have not been evaluated for low cell input from multicellular systems. Internal RNA target labeling has been recently 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]. TRIBE allowed detection of RBP targets using as little as 150 *Drosophila* neurons, and the approach has since been updated using an ADAR hyperactivating mutation to increase editing efficiency (HyperTRIBE) [18-22]. Even with this hyperactivating mutation, these ADAR-mediated approaches are constrained by the sparsity of double-stranded regions proximal to RBP binding sites that are required for ADAR-mediated adenosine to inosine editing [23].

APOBEC1 is a cytosine deaminase that catalyzes RNA cytosine-to-uracil (C-to-U) conversion [24]. Recently APOBEC1 was fused to the m6A-binding YTH domain to identify m6A modification sites on RNAs genome-wide [25]. This approach, termed DART-seq (Deamination Adjacent to RNA modification Target), is distinguished from TRIBE as it departs from using ADAR dsRNA A-to-I editing domains. DART-seq showed precision for identifying m6A-YTH interactions on single stranded mRNAs and demonstrated binding site

resolution with good specificity. However, it was unclear if APOBEC1 fusions would work with non-m6A or general classes of RBPs such as splicing factors, RNA stability factors, histone stem-loop binding proteins, or even ribosomes. We reasoned that fusion of APOBEC1 to such full-length RBPs (rather than a section of an RBP such as the YTH domain) could allow for robust, immunoprecipitation-free identification of RBP targets across functional RNA-interaction categories using extremely low or even single-cell input. Here, we demonstrate the utility 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 greatly extends the DART-seq approach to demonstrate, for the first time, the discovery of RBP-RNA sites at single-cell resolution, the deconvolution of targets for multiplexed RBPs, and the cell-type specific binding of an RBP in a heterogenous mixtures of cell-types. By applying STAMP with specific ribosome subunits, we also extend this approach for single-cell detection of ribosome association while simultaneously measuring gene expression. Our innovations demonstrate that translation efficiency and RBP-interactomes can be measured at single-cell level, opening up new paradigms for probing biological questions.

Results

STAMP identifies RBP binding sites without immunoprecipitation

Our strategy for immunoprecipitation-free detection of RBP targets involves fusing an RBP of interest to the cytidine deaminase enzyme APOBEC1, which is known to catalyze C-to-U editing on single-stranded RNA targets (**Figure 1A**). We termed this general strategy STAMP (Surveying Targets by APOBEC Mediated Profiling). Upon expression of an RBP-APOBEC1 fusion protein (RBP-STAMP), RBPs direct the deaminase module to specific RNA targets leading to C-to-U base conversion proximal to RBP binding sites. These mutations are resolved using high-throughput RNA sequencing approaches and quantified using the SAILOR analysis pipeline [26], which we modified to identify and assign a confidence score for C-to-U mismatches using a beta distribution that factors both site coverage and editing percentage (see Materials and Methods) following removal of annotated hg19 SNPs [27].

To determine the utility of the STAMP approach, we fused APOBEC1 to the C-terminus of the RBP RBFOX2 [28-30] 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 high (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) [31] (**Figure 1B**). Uninduced RBFOX2-STAMP, or control-STAMP (APOBEC1 only) at low and high induction, had few or no detectible 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 confidence levels, respectively, as called by SAILOR (**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. All RBFOX2-STAMP induction levels and time points showed reproducible edit enrichment on the *APP* 3'UTR eCLIP peak region (**Figure S1A**), when compared to undetectable edits from corresponding control-STAMP (**Figure S1B**). We found that the number of edits (confidence ≥ 0.5) on a given target gene was highly correlated between replicates, and that transcriptome-wide correlations improved with higher RBFOX2-STAMP levels (**Figure 1D, Figure S1C**). Irreproducible discovery rate (IDR) analysis [32] revealed hundreds to thousands of reproducible edit clusters for RBFOX2-STAMP, with the number of these reproducible clusters also increasing with levels of RBFOX2-STAMP (**Figure S1D**). 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 24, 48 and 72 hours compared to uninduced controls and detected negligible changes in gene expression at these collection points (**Figure S1E, Table S3**).

We next measured the nucleotide distance of RBFOX2-STAMP edits from known RBFOX2 binding sites. For the 2,852 RBFOX2 eCLIP peaks that harbor the canonical RBFOX2 motif GCAUG, distances from the motif to RBFOX2-STAMP and control-STAMP (background) edits were determined within a 400 bp window (**Figure 1E**). We observed 1,751 edits for RBFOX2-STAMP within 200 bp of binding site motifs inside eCLIP peaks, compared to 97 edits from control-STAMP, indicating that RBFOX2 RNA-binding activity is directing and enriching RBFOX2-specific edits at conserved sites. This RBFOX2-STAMP edit enrichment decreased to background >200 bp away from the motif. We next measured the number of high confidence (≥ 0.99) edits in 100-bp windows centered around RBFOX2 eCLIP peaks within coding sequence (CDS) and 3'UTRs. We saw that ~30% of eCLIP peaks contained at least one high confidence edit, and importantly, the ratio of fractions comparing authentic peaks to randomly located sequences within the same annotated regions (n trials = 100) increased with the number of edits found in each window (blue line in **Figure 1F**). To establish a metric for reporting STAMP signal across regions and whole genes that takes into account (i) read coverage over edit site, (ii) edit frequency and (iii) editing potential (C-encoded reference bases), we calculated an edit fraction score (ε) as the ratio of the total SAILOR-filtered C-to-U reads over the total SAILOR-filtered mapped reads across every annotated C within a given sequence window. Binning windows evenly across the range of ε scores revealed that this score is a strong indicator of edit enrichment (blue line) within eCLIP peaks compared to randomly located regions (**Figure 1G**). Finally, we performed *de novo* motif discovery within merged 51-nucleotide windows filtered only by confidence score, assessing enrichment above a shuffled background for each gene region. Sequences in regions containing high confidence (≥ 0.99) RBFOX2-STAMP edits 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 1H**). Importantly, no level of control-STAMP expression at any window size or confidence level derived the (U)GCAUG motif (**Figure 1H**) and we noted that comparison of all detectible edits from high control- and RBFOX2-STAMP inductions resulted in only a 3.3% overlap in edit sites, demonstrating the sensitivity and specificity of STAMP for determining RBP binding sites.

STAMP identifies known RNA targets in multiple RBP backgrounds

Next we 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, 33, 34]. SLBP has a very specific set of characterized RNA targets, binding histone mRNAs at 20 nucleotide stem loop regions near the very 3' end of 3'UTRs preceding non-polyadenylated cleavage sites [33]. Comparison of SLBP-STAMP to SLBP eCLIP data [35] showed that SLBP-STAMP edits were reproducibly enriched compared to control-STAMP in close proximity (within 50 nucleotides upstream or downstream) from replicate SLBP eCLIP peaks within the 3'UTR of histone genes, such as *H1-2* (**Figure 2A**) and *H4C12* (**Figure 2B**) adjacent to stem loop regions, as expected. We calculated ε scores for SLBP-STAMP on SLBP eCLIP peaks within histone CDS and 3'UTR regions and saw that the fraction of SLBP eCLIP peaks that contained high ε scores increased relative to the fraction of randomly located regions (**Figure 2C**). Comparison of either merged or replicate SLBP-STAMP ε scores to RBFOX2-STAMP ε scores across all 12,149 genes edited by either RBP revealed RBP-specific edit enrichment (RBFOX2 eCLIP targets in blue, SLBP histone gene targets in magenta), further demonstrating that STAMP editing reflected the binding specificity of the fused RBP (**Figure 2D, Figure S2A**).

Much like RBFOX2- and SLBP-STAMP, comparison of control- and TIA1-STAMP to TIA1 eCLIP revealed that there was reproducible C-to-U edit enrichment arising from low and high level TIA1-STAMP on a significantly enriched eCLIP 3'UTR peak within the *TMCO1* gene (**Figure 2E**). Transcriptome-wide, the TIA1-STAMP edit fraction was enriched around eCLIP peaks based upon ε scores (**Figure 2F**), and high confidence edits ($\geq$0.999) were able to extract the known eCLIP established U(A)-rich binding motif specific for TIA1 at high doxycycline induction (**Figure 2G**). 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 (**Figure S2B**), and IDR analysis revealed thousands of reproducible edit clusters that increased in number with increasing induction levels (**Figure S2C**). 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 had previously observed that eCLIP of small ribosomal subunit RPS3 featured binding patterns that recapitulated the average profile captured by ribosome profiling [3]. Here, we generated independent HEK293T cell lines expressing APOBEC1 fusions to ribosomal subunits RPS2 and RPS3. For RPS2-STAMP and RPS3-STAMP we observed that C-to-U edits were enriched relative to control-STAMP on exons of protein-coding genes that are highly translated in HEK293T cells, such as *ATP5BP* [36], coincident with RPS3 eCLIP

signal enrichment over size-matched input control (**Figure 3A**). In comparison, RPS2-STAMP and RPS3-STAMP signal were minimally detected on non-coding genes such as the lncRNA *MALAT1*, which is localized to the cytoplasm in mitotic cell lines [37](**Figure 3B**). We performed replicate RPS2-STAMP and control-STAMP inductions at low and high doxycycline concentrations for 24, 48 and 72 hours and observed high, dose-dependent edit count reproducibility, as well as low overlap (2.8% of all detectable edits) between control-STAMP and RPS2-STAMP edit sites at high induction (**Figure S3A, Table S4**).

To evaluate whether Ribo-STAMP can distinguish genes with varying levels of ribosome occupancy, we compared to published ribo-seq, RiboLace [10] and polysome profiling data [38] obtained from HEK293 cells. We grouped transcripts into ribosome occupancy quartiles as measured by these independent methods and quantified the distribution of ε scores from RPS2- and RPS3-STAMP compared to control-STAMP for genes in each quartile. Absent induction of RPS2-STAMP, ε score distributions between RPS2- and control-STAMP were indistinguishable (**Figure 3C, left**). At low levels of induction, we found that ε scores from RPS2-STAMP were significantly higher than control-STAMP, particularly in upper ribosome occupancy quartiles (**Figure 3C, middle**). These statistically significant differences (Q1, Q2; $p < 1e\text{-}300$, Q3 $p = 1e\text{-}227$, Wilcoxon rank-sum) were accentuated at high induction of RPS2- and control-STAMP conditions (**Figure 3C, right,** Q1-Q3; $p < 1e\text{-}300$, Wilcoxon rank-sum). We observed the same effect when comparing RPS2-STAMP to RiboLace (**Figure 3D**) and polysome profiling (**Figure 3E**). Because RPS2-STAMP had more consistent signal across CDS regions and higher correlation with independent ribosome foot-printing approaches than RPS3-STAMP (**Figure 3F**), 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 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 3G**)**.** Enrichment of RPS2-STAMP edits within 3'UTRs likely indicates ribosome association with these accessible regions following ribosome translation termination by release factors, as we also observed 3'UTR signal from RPS3 eCLIP (**Figure 3A**).

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 [39]. Pharmacological inhibition of the mTOR pathway globally suppresses translation of mRNAs after initially suppressing translation of genes encoding the translational machinery itself [40]. 48-hour Torin-1 treatment resulted in reproducible suppression in RPS2-STAMP edit distributions compared to untreated cells, exemplified by a marked decrease specifically in coding sequence edits on the top quartile of ribosome occupied genes (ribo-seq, **Figure 3H**). RPS2-STAMP ε scores were also significantly reduced upon 48-hour Torin-1 treatment in the highest quartile of ribosome occupied genes as defined by ribo-seq (Q1 $p = 3.9\ e\text{-}98$, Wilcoxon rank-sum), RiboLace (Q1 $p = 8.4\ e\text{-}87$, Wilcoxon rank-sum) and polysome profiling (Q1 $p = 1.2\ e\text{-}54$, Wilcoxon rank-sum), and all reported Torin-1 sensitive TOP genes [40] were contained within these top quartiles (**Figure 3H**). We observed no significant differences in ε scores for control-STAMP cells upon 48-hour Torin-1 treatment for any matched comparisons

(**Figure S3B**). Together these results demonstrate that specific and dynamic translational responses are detectible by Ribo-STAMP.

Long-read STAMP reveals isoform specific binding profiles.

Given that STAMP does not require isolation of RBP-protected RNA fragments, unlike commonly performed CLIP assays, we hypothesized that STAMP may enable 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 both RNA and cDNA long reads with the Oxford Nanopore sequencing platform [41, 42]**.** Both of these direct long-read sequencing approaches resulted in enrichment above control of C-to-U edits from RBFOX2-STAMP that overlapped with both eCLIP signal and short read (Illumina) RBFOX2-STAMP signal on the target gene *APP* 3'UTR (**Figure 4A**).

Long-read high confidence (≥ 0.99) RBFOX2-STAMP edits extracted the known RBFOX2 UGCAUG binding motif using either genomic mapping (hg19; p = 1e-42, hypergeometric distribution) or isoform mapping (cDNA; p = 1e-27, hypergeometric distribution). Encouraged by the ability to detect RBFOX2 edits on cDNA alignments we wanted to evaluate if there were any isoform-specific binding events. To accomplish this, we counted RBFOX2-STAMP or control-STAMP edits above confidence level 0.99 on the primary and secondary isoforms of 1,631 RBFOX2 eCLIP target genes, and observed differential isoform editing signatures (**Figure 4B**). To illustrate, displayed edits on the *UNK* (**Figure 4C, D**) and *IDS* **(Figure S4A, B)** genes and observed RBFOX2-STAMP (but not control-STAMP) isoform specific edits, implying isoform-specific binding by RBFOX2 to these transcripts. 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 leveraged a commercially available single-cell capture platform. 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 854 true RBFOX2-STAMP cells to compare to 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 5A**). Both the number of edited genes (**Figure S5A**) and the summed ε score per cell (**Figure S5B**) were higher in RBFOX2-STAMP cells compared to control-STAMP cells, and single-cell RBFOX2-STAMP edits were detected on 71% of genes that also contained bulk RBFOX2-STAMP edits, with only 7% of single-cell edited genes not detected in bulk (**Figure S5C**)**.** For normalized edit

counts across all expressed gene 3'UTRs (n = 15,832 genes, confidence level ≥ 0.9), bulk and single-cell RBFOX2-STAMP showed strong correlation (**Figure 5B,** R^2 = 0.61, p < 1e-300, two-sided test with R beta distribution). 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 clusters from both bulk RBFOX2-STAMP and the aggregate of all RBFOX2-STAMP cells (**Figure 5C**), revealing that STAMP can define RBP binding sites at single-cell resolution. *De novo* motif analysis from randomly down-sampled numbers of cells identified the canonical (U)GCAUG motif with significance, even to the resolution of one cell (**Figure 5D**) 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 could 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 far better distinguished by UMAP visualization using ε scores, than by gene expression (**Figure 6A**, **S6A**), 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 TIA-STAMP "capture cells" (**Figure 6B**). Using Louvain clustering by ε score profiles we thus defined an RBFOX2-cluster (n = 6,003 cells), a TIA1-cluster (n = 1,841 cells) and a "background"-cluster (n = 6,623 cells) for further analysis (**Figure S6B**). Overlap with control (**Figure 6C**) and re-clustering in the expression space (**Figure S6C**) 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 S6D**), and TIA1-and RBFOX2-clusters displayed distinct editing profiles when compared to control-STAMP (**Figure 6D, 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*. Individual cell edit enrichments were specific to TIA1-STAMP on the *BTF3* target gene, and to RBFOX2-STAMP on the *CFL1* target gene (**Figure 6E**), demonstrating that the targets and binding sites of multiplexed RBP-STAMP fusions can be delineated from edit signatures within single-cell experiments.

We next set out 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) [43] 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 6F, Figure S6E, 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 6G, 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 S6F**). Analysis of the top RBFOX2-STAMP differentially edited genes between cell types revealed cell-type specific targets (**Figure 6H**) that were often not differentially expressed (**Figure 6I**), 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 6J**). 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. 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 S7A**), 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: (i) RPS2-cluster cells (n = 3,868 cells), (ii) RBFOX2-cluster cells (n = 7,000 cells), containing the majority (92%) of RBFOX2 cells identified by capture sequencing, (iii) TIA1-cluster cells (n = 1312 cells) containing the majority (57%) of TIA1 capture cells, and (iv) a background-cluster population (n = 20,655 cells), composed of control-STAMP cells and any cells that overlap spatially with control-STAMP cells (**Figure 7A and S7B**). While UMAP visualization of gene expression could distinguish these individual STAMP experiments (**Figure S7C**), expression profiles were insufficient for determining background cells (**Figures S7D**), again highlighting the importance of ε score clustering and control-STAMP overlap for defining specific RBP cell groups.

For the RPS2-cluster cells, we observed a significant correlation (R^2 = 0.48, p < 1e-300, two-sided test with R beta distribution) in edit counts on 15,044 genes between bulk RPS2-STAMP and single-cell aggregate (**Figure 7B**). As we saw for bulk RPS2-STAMP ε scores, single-cell RPS2-STAMP ε scores were higher on genes with higher ribosome occupancy as detected by ribo-seq (**Figure 7C**), indicating that single-cell Ribo-STAMP, like single-cell RBP-STAMP, recapitulates results from bulk experiments and can define single-cell ribosome association.

Differential ε score analysis showed distinct editing signatures for RPS2-, RBFOX2- and TIA1-cluster cells compared to the background-cluster (**Figure 7D, Table S8**). 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 7E**), in agreement with our results from bulk (**Figure 3G**). 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 7F**). These results highlight the capability of STAMP to simultaneously reveal RBP targets and ribosome association at single-cell resolution.

To determine if Ribo-STAMP can be used to detect different translational landscapes between cells, we focused on RPS2-STAMP versus control-STAMP cells (**Figure S7E**) and used Louvain clustering to define background- and RPS2-clusters **(Figure S7F**). We then subdivided the RPS2-cluster population by ε score Louvain clustering into six distinct subpopulations (**Figure 7G**). Given that we detected RPS2-STAMP expression from all RPS2-cluster subsets within the same cell type (**Figure 7H**), we reasoned that this editing heterogeneity may arise from differential translation within different phases of the mitotic cell cycle. To test this hypothesis, we assigned RPS2-cluster cells into the G1, S and G2M phases of the cell cycle based on gene expression [44] (**Figure S7G, Table S9**). Gene ontology analysis of genes edited specifically in G2M phase yielded enriched terms related to M phase (**Figure S7H**). Specific RPS2-cluster subsets were associated with specific cell cycle edit signatures. In particular, RPS2-cluster subsets 1 (n = 961 cells) and 5 (n = 458 cells) shared editing signatures with G2M phase cells (**Figure 7I, Table S10**), and the differential editing of these cell cycle genes was not simply explained by differential expression (**Figure S7I**). To discern cell cycle translation signatures from transcription signatures, we compared RPS2-STAMP ε scores to cell cycle specific translation efficiency (TE) measurements previously determined by ribo-seq in M phase-arrested cells [45]. We observed enriched edits on transcripts with high M phase TE (ribo-seq, Park et al. gene-set 2 [45]) in cells from RPS2-cluster subsets 1 and 5, which were predicted to be in M phase by our expression analysis (**Figure 7J**). These results demonstrate the ability of Ribo-STAMP to detect ribosome associations that reflect specific translational signatures at single-cell resolution.

Discussion

We have developed STAMP and its companion computational workflow as a framework 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. The recently published DART-seq approach was used to mark m6A RNA modifications genome wide [25], however it was unclear if the method would work for general classes of full-length RBPs and even ribosomes. To distinguish our framework, which showcases unprecedented single-cell resolution binding sites of a broad range of RBPs and ribosome subunits, from DART-seq (a portion of the YTH-domain) and TRIBE (ADAR deaminase domains), we call our framework STAMP (Surveying Targets by Apobec-mediated Profiling). We were able to demonstrate the specificity of STAMP for diverse, full-length RBPs that bind both polyadenylated mRNAs (RBFOX2, TIA1) and non-polyadenylated mRNAs (SLBP). We also demonstrate that ribosomal subunits RPS2 and RPS3 when fused to APOBEC1 enable the measurement of ribosome association that correlates well with metrics such as translation efficiency computed from ribo-seq, RiboLace and polysome profiling experiments which often require separate mRNA-seq data as normalizing denominators. 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. Furthermore, with single-cell Ribo-STAMP we observed that cell populations with gene expression signatures characteristic of the M phase of the mitotic cell cycle were enriched for edits on M phase translated genes. This was consistent with data by standard ribosome profiling in M phase arrested cells [45-47], highlighting the potential of single-cell Ribo-STAMP to simultaneously examine both transcriptional and translational levels of gene expression control. 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, 19, 21, 22]. 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 [23], an infrequent occurrence on single stranded mRNAs transcriptome-wide [18, 23, 48]. 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). We observe this density of edits with full-length RBPs, but it is apparent even with only the YTH domain of the m6a reader YTHDF2 [25]. In addition, structured RNA is reduced from coding regions by active translation [49], 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 [18]. 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 transcripts in 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, with eCLIP-seq as a comparison, we are constrained by the input requirements, frequently needing thousands to millions of cells for a successful eCLIP experiment. 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. As a result, even without single cell analyses, STAMP is permissive of ultra-low input material which enables rare cell-types to be collected and analyzed for RBP-interactomes. 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 [50], which opens the possibility of using STAMP to detect modification-sensitive RNA-protein interactions.

Caveats and considerations

Current versions of STAMP have several limitations. 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 roles mediated by interacting with sites in mRNA, RBFOX2 and TIA1 are also splicing factors with binding sites within intronic regions which are missed 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. On the other hand, false positive binding sites are also possible when expression levels of STAMP transgenes are supra-physiological, leading to promiscuous RNA interactions. Expression of fusions from a native promoter could help address this concern. Certainly, RBP-STAMP signal that does not overlap with traditional CLIP peaks may indeed be *bona fide* binding events that CLIP methodologies are not sensitive enough to detect. Unlike RBFOX2, SLBP and TIA1 that have conserved and narrow binding sites, other RBPs may exhibit broad binding patterns across the sequence of the transcripts, such as the ribosomal subunits RPS2 and RPS3.

Our current version of Ribo-STAMP yields detectable edits within 12-24 hours. However, we do observe gene expression changes on a subset of genes at high Ribo-STAMP induction after 72 hours, likely 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 may lead to downstream unintended perturbations. If longer time frames are needed, use of STAMP with the ribosome associated factors, such as translation initiation or release factors, may allow detection of translation responses without long-term coding sequence perturbations. Alternatively, we can develop more rapid and responsive ribosome directed edits to achieve snapshots of translatome responses on shorter timescales. 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 [18, 20, 21, 25]. EIF4E-BP-HyperTRIBE, which sought to use A-to-I editing to mark 5' ends of translationally repressed transcripts, observed a similar 3'UTR enrichment to Ribo-STAMP, which was potentially explained by 5' and 3' end association of translation components and mRNA features [20]. In the future, we anticipate that engineering of fusion orientation, altering STAMP expression levels and duration of overexpression window may be useful to obtain editing profiles that are maximally informative for different RBPs.

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. Previously, cell type-specific CLIP or Ribo-seq from in vivo models has remained a challenge in the field. The STAMP approach, utilizing cell-type specific promoters and gene expression profiles, would overcome these major challenges. Organoid or animal 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.

Acknowledgements

We thank Thai Nguyen for generating TIA1 eCLIP datasets. We thank Dr. Gabriella Viero at the Institute of Biophysics, CNR Unit at Trento for her helpful advice concerning analysis of ribo-seq and RiboLace data. We thank Dr. Todd Michael at the J. Craig Venter Institute for use of the Oxford Nanopore PromethION system. We thank 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 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. 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.

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, 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, 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.

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.
2. Gerstberger, S., M. Hafner, and T. Tuschl, *A census of human RNA-binding proteins.* Nat Rev Genet, 2014. **15**(12): p. 829-45.
3. Van Nostrand, E.L., et al., *Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins.* Genome Biol, 2020. **21**(1): p. 90.

4. Martinez, F.J., et al., *Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System.* Neuron, 2016. **92**(4): p. 780-795.
5. Ramanathan, M., D.F. Porter, and P.A. Khavari, *Methods to study RNA-protein interactions.* Nat Methods, 2019. **16**(3): p. 225-234.
6. Perez-Perri, J.I., 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.
7. Calviello, L. and U. Ohler, *Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome.* Trends Genet, 2017. **33**(10): p. 728-744.
8. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.* Science, 2009. **324**(5924): p. 218-23.
9. Lee, F.C.Y. and J. Ule, *Advances in CLIP Technologies for Studies of Protein-RNA Interactions.* Mol Cell, 2018. **69**(3): p. 354-369.
10. Clamer, M., et al., *Active Ribosome Profiling with RiboLace.* Cell Rep, 2018. **25**(4): p. 1097-1108 e5.
11. Buenrostro, J.D., 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.
12. Hwang, B., J.H. Lee, and D. Bang, *Single-cell RNA sequencing technologies and bioinformatics pipelines.* Exp Mol Med, 2018. **50**(8): p. 96.
13. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell.* Nat Methods, 2009. **6**(5): p. 377-82.
14. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nat Methods, 2017. **14**(9): p. 865-868.
15. Shahi, P., et al., *Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding.* Sci Rep, 2017. **7**: p. 44447.
16. Lapointe, C.P., et al., *Protein-RNA networks revealed through covalent RNA marks.* Nat Methods, 2015. **12**(12): p. 1163-70.
17. Medina-Munoz, H.C., et al., *Records of RNA locations in living yeast revealed through covalent marks.* Proc Natl Acad Sci U S A, 2020.
18. McMahon, A.C., et al., *TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins.* Cell, 2016. **165**(3): p. 742-53.
19. Nguyen, D.T.T., et al., *HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells.* Nat Commun, 2020. **11**(1): p. 2026.
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. Xu, W., R. Rahman, and M. Rosbash, *Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein.* RNA, 2018. **24**(2): p. 173-182.
22. Rahman, R., et al., *Identification of RNA-binding protein targets with HyperTRIBE.* Nat Protoc, 2018. **13**(8): p. 1829-1849.
23. Matthews, M.M., 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.
24. 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.
25. Meyer, K.D., *DART-seq: an antibody-free method for global m(6)A detection.* Nat Methods, 2019. **16**(12): p. 1275-1280.
26. Deffit, S.N., et al., *The C. elegans neural editome reveals an ADAR target mRNA required for proper chemotaxis.* Elife, 2017. **6**.
27. Washburn, M.C., 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.
28. Lovci, M.T., et al., *Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges.* Nat Struct Mol Biol, 2013. **20**(12): p. 1434-42.

29. Yeo, G.W., 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.
30. Ponthier, J.L., 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.
31. Van Nostrand, E.L., et al., *CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins.* Methods, 2017. **118-119**: p. 50-59.
32. Li, Q.H., et al., *Measuring Reproducibility of High-Throughput Experiments.* Annals of Applied Statistics, 2011. **5**(3): p. 1752-1779.
33. Marzluff, W.F., E.J. Wagner, and R.J. Duronio, *Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail.* Nat Rev Genet, 2008. **9**(11): p. 843-54.
34. 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.
35. Davis, C.A., et al., *The Encyclopedia of DNA elements (ENCODE): data portal update.* Nucleic Acids Res, 2018. **46**(D1): p. D794-D801.
36. Li, B.B., 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.
37. Yang, F., et al., *MALAT-1 interacts with hnRNP C in cell cycle regulation.* FEBS Lett, 2013. **587**(19): p. 3175-81.
38. Tan, F.E., et al., *A Transcriptome-wide Translational Program Defined by LIN28B Expression Level.* Mol Cell, 2019. **73**(2): p. 304-313 e3.
39. Thoreen, C.C., 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.
40. Thoreen, C.C., et al., *A unifying model for mTORC1-mediated regulation of mRNA translation.* Nature, 2012. **485**(7396): p. 109-13.
41. Jain, M., et al., *The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community.* Genome Biol, 2016. **17**(1): p. 239.
42. Branton, D., et al., *The potential and challenges of nanopore sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1146-53.
43. Li, Y., et al., *A comprehensive library of familial human amyotrophic lateral sclerosis induced pluripotent stem cells.* PLoS One, 2015. **10**(3): p. e0118266.
44. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
45. Park, J.E., et al., *Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle.* Mol Cell, 2016. **62**(3): p. 462-471.
46. Tanenbaum, M.E., et al., *Regulation of mRNA translation during mitosis.* Elife, 2015. **4**.
47. Stumpf, C.R., et al., *The translational landscape of the mammalian cell cycle.* Mol Cell, 2013. **52**(4): p. 574-82.
48. Song, Y., et al., *irCLASH reveals RNA substrates recognized by human ADARs.* Nat Struct Mol Biol, 2020. **27**(4): p. 351-362.
49. Beaudoin, J.D., et al., *Analyses of mRNA structure dynamics identify embryonic gene regulatory programs.* Nat Struct Mol Biol, 2018. **25**(8): p. 677-686.
50. Lorenz, D.A., 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.

Figures and Legends:

Figure 1

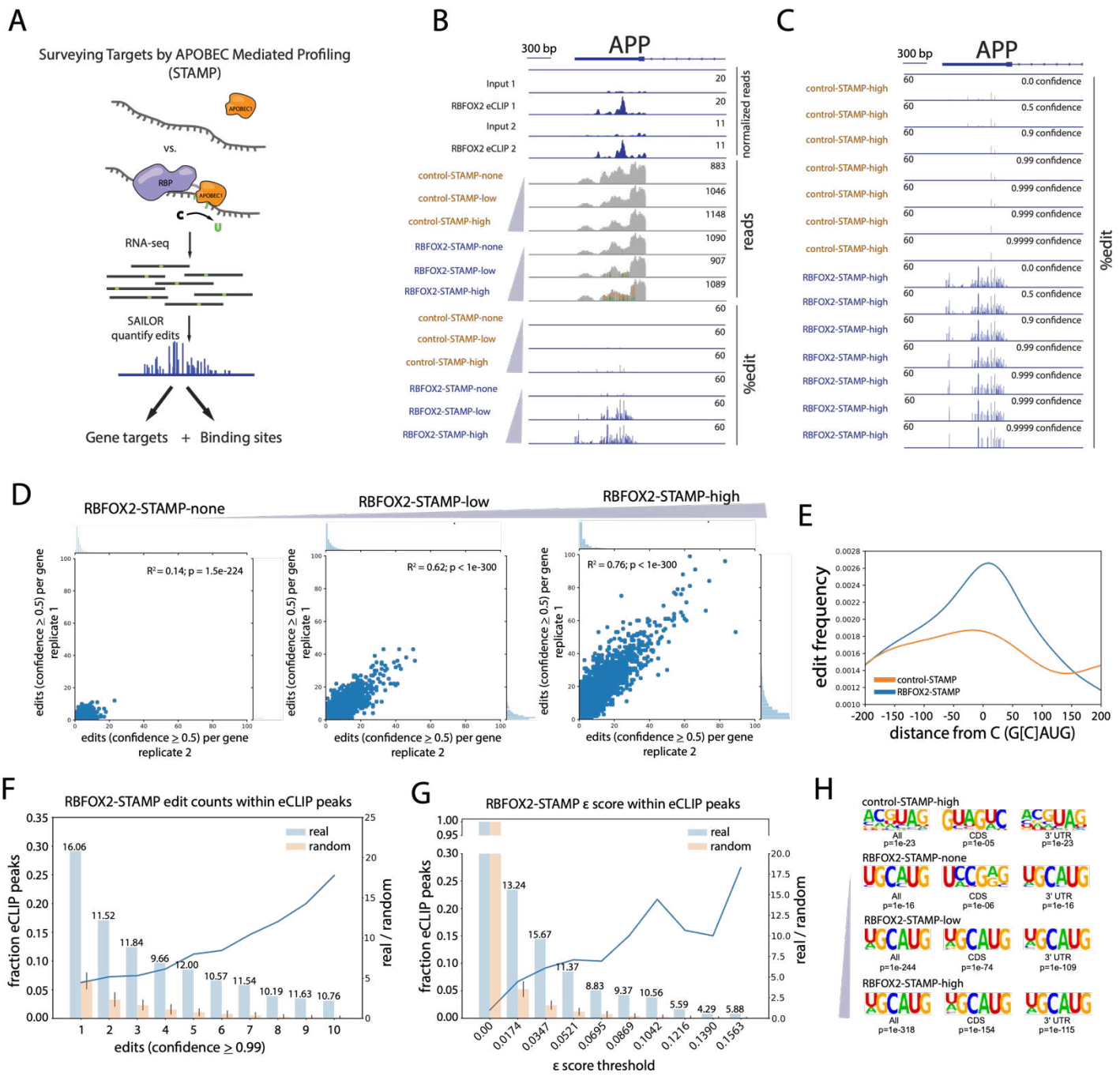

A
Surveying Targets by APOBEC Mediated Profiling (STAMP)
vs.
RBP
APOBEC1
RNA-seq
SAILOR quantify edits
Gene targets + Binding sites
B
300 bp
APP
Input 1
RBFOX2 eCLIP 1
Input 2
RBFOX2 eCLIP 2
normalized reads
control-STAMP-none
control-STAMP-low
control-STAMP-high
RBFOX2-STAMP-none
RBFOX2-STAMP-low
RBFOX2-STAMP-high
reads
%edit
C
300 bp
APP
0.0 confidence
0.5 confidence
0.9 confidence
0.99 confidence
0.999 confidence
0.9999 confidence
%edit
D
RBFOX2-STAMP-none
RBFOX2-STAMP-low
RBFOX2-STAMP-high
R² = 0.14; p = 1.5e-224
R² = 0.62; p < 1e-300
R² = 0.76; p < 1e-300
edits (confidence ≥ 0.5) per gene replicate 1
edits (confidence ≥ 0.5) per gene replicate 2
E
edit frequency
control-STAMP
RBFOX2-STAMP
distance from C (G[C]AUG)
F
RBFOX2-STAMP edit counts within eCLIP peaks
real
random
fraction eCLIP peaks
real / random
edits (confidence ≥ 0.99)
G
RBFOX2-STAMP ε score within eCLIP peaks
real
random
fraction eCLIP peaks
real / random
ε score threshold
H
control-STAMP-high
RBFOX2-STAMP-none
RBFOX2-STAMP-low
RBFOX2-STAMP-high
All
CDS
3' UTR

Figure 1: RBFOX2-STAMP enriches C-to-U edits at RBFOX2 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 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 number of edits (confidence level ≧ 0.5) per target (p values assessed with two-sided test with R beta distribution) **E**) Edit frequency distribution within a 400 bp window flanking RBFOX2 binding-site motifs for RBFOX2-STAMP and control-STAMP (background) **F**) Fraction of RBFOX2 motif-containing eCLIP peaks in 3'UTR and CDS regions with increasing numbers of ≧ 0.99 confidence edits compared to size-matched randomized regions outside of eCLIP peaks, z-scores reported above bars **G**) Fraction of RBFOX2 motif-containing eCLIP peaks or random regions, as in G, with increasing edit fraction score (ε score) thresholds **H**) Motif enrichment using HOMER and shuffled background on combined control- or RBFOX2-STAMP ≥ 0.99 confidence level edit windows for increasing RBFOX2-STAMP induction levels.

Figure 2

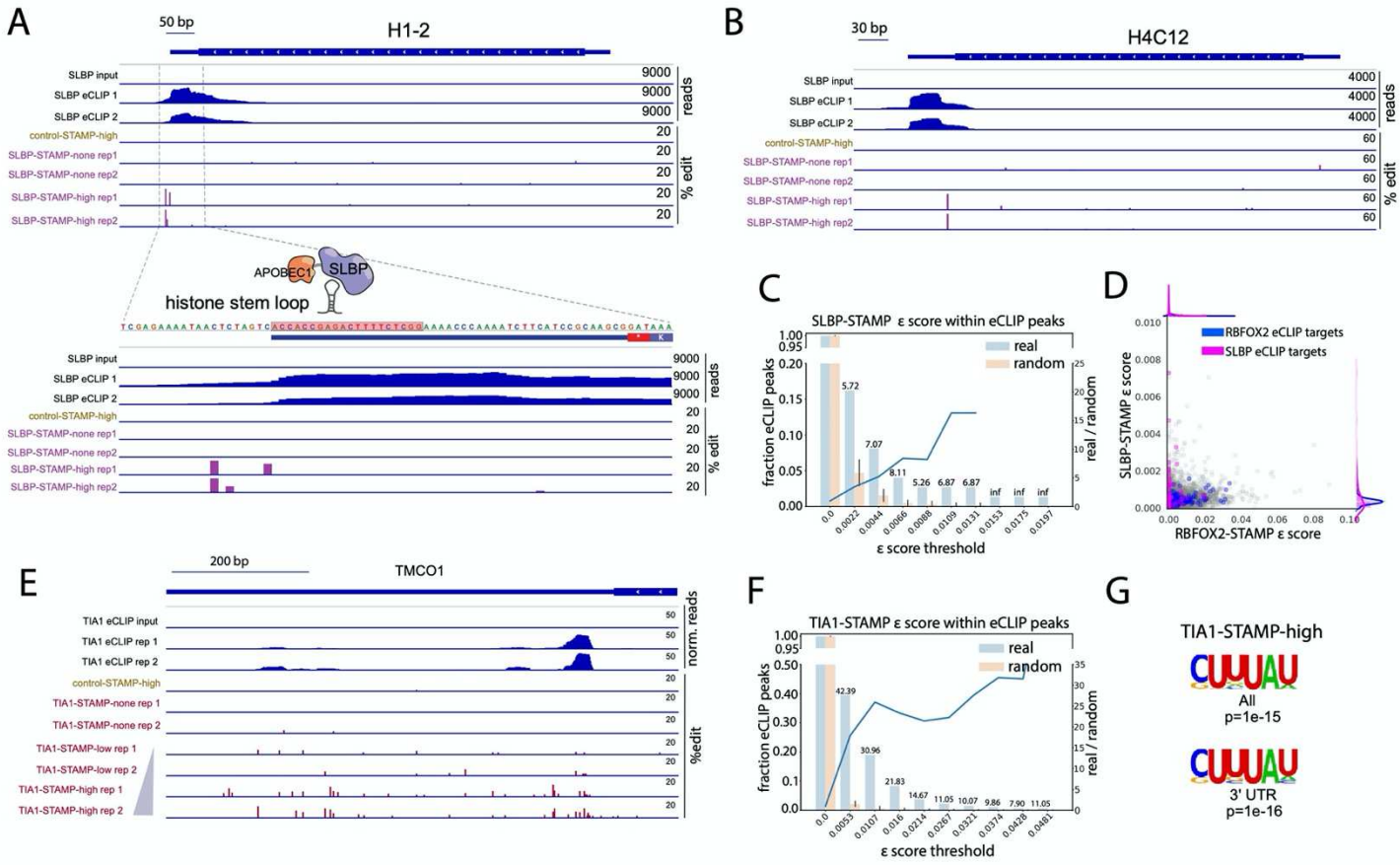

A
50 bp
H1-2
SLBP input
SLBP eCLIP 1
SLBP eCLIP 2
control-STAMP-high
SLBP-STAMP-none rep1
SLBP-STAMP-none rep2
SLBP-STAMP-high rep1
SLBP-STAMP-high rep2
reads
% edit
APOBEC1
SLBP
histone stem loop
B
30 bp
H4C12
C
SLBP-STAMP ε score within eCLIP peaks
real
random
fraction eCLIP peaks
real / random
ε score threshold
D
RBFOX2 eCLIP targets
SLBP eCLIP targets
SLBP-STAMP ε score
RBFOX2-STAMP ε score
E
200 bp
TMCO1
TIA1 eCLIP input
TIA1 eCLIP rep 1
TIA1 eCLIP rep 2
control-STAMP-high
TIA1-STAMP-none rep 1
TIA1-STAMP-none rep 2
TIA1-STAMP-low rep 1
TIA1-STAMP-low rep 2
TIA1-STAMP-high rep 1
TIA1-STAMP-high rep 2
norm. reads
%edit
F
TIA1-STAMP ε score within eCLIP peaks
G
TIA1-STAMP-high
All
p=1e-15
3' UTR
p=1e-16

Figure 2: **SLBP-STAMP and TIA1-STAMP expression enriches edits on respective targets**. **A**) IGV tracks showing replicate SLBP-STAMP edit fractions at no- and high-induction (doxycycline: 0ng = none or 1µg/ml = high, 72 hours) on the target histone gene *H1-2* compared to SLBP eCLIP, highlighting SLBP-bound histone stem loop **B**) IGV tracks as in A showing replicate SLBP-STAMP edit fractions on the target histone gene *H4C12* compared to SLBP eCLIP **C**) Fraction of SLBP eCLIP peaks or random regions with increasing ε score thresholds, z-scores reported above bars **D**) Scatterplot comparing ε score for merged replicate high-induction SLBP-STAMP with histone gene targets (magenta, n = 89) highlighted compared to high-induction RBFOX2-STAMP ε score and RBFOX2 eCLIP peak containing genes (blue, n = 342) highlighted **E**) browser tracks showing TIA1 eCLIP peaks on the target gene *TMCO1*, compared with replicate TIA1-STAMP edit fraction for increasing induction levels of TIA1-STAMP **F**) Fraction of TIA1 eCLIP peaks or random regions with increasing ε score thresholds, z-scores reported above bars **G**) Motif enrichment using HOMER and shuffled background on combined TIA1-STAMP ≥ 0.999 confidence level edit windows for high TIA1-STAMP induction levels

Figure 3

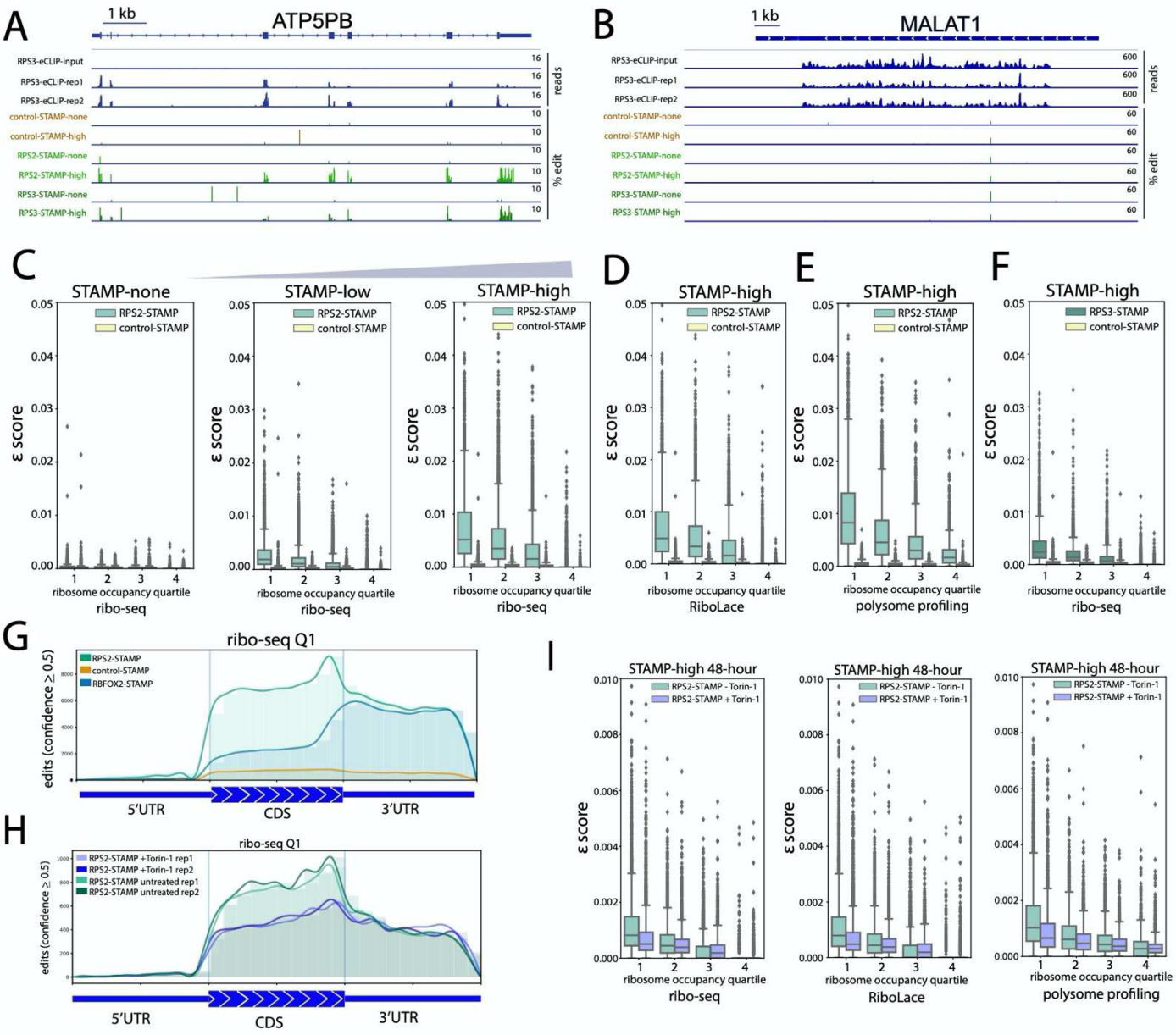


Figure 3: **Ribo-STAMP edits mark highly translated coding sequences. A**) IGV browser tracks displaying coding sequence edit frequency from control, RPS2-STAMP, and RPS3-STAMP at no-induction or 72-hour high-

induction on the *ATP5BP* 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 or RPS2-, RPS3-STAMP. **C**) Comparison of gene quartiles ranked by ribosome occupancy (ribo-seq) with ε score from increasing levels of RPS2-STAMP and control-STAMP (Q1, Q2 $p < 1e\text{-}300$ for low and high RPS2-STAMP, Wilcoxon rank-sum) **D**) Comparison of gene quartiles ranked by ribosome occupancy (RiboLace) with ε score from high induction of RPS2-STAMP (Q1-Q3 $p < 1e\text{-}300$, Wilcoxon rank-sum) **E**) Comparison of gene quartiles ranked by ribosome occupancy (polysome profiling) with ε score from high induction of RPS2-STAMP (Q1-Q4 $p < 1e\text{-}300$, Wilcoxon rank-sum) **F**) Comparison of gene quartiles ranked by ribosome occupancy (ribo-seq) with ε score from high induction of RPS3-STAMP (Q1,Q2 $p < 1e\text{-}300$, Wilcoxon rank-sum) **G**) Metagene plot showing edit (≥ 0.5 confidence level) 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) **H**) Metagene plot showing edit (≥ 0.5 confidence level) distribution for untreated 48-hour high-induction RPS2-STAMP compared to replicate Torin-1 treated 48-hour high-induction RPS2-STAMP across 5'UTR, CDS and 3'UTR gene regions for the top quartile of ribosome occupied genes (ribo-seq, n = 4931 genes) **I**) Comparison of ε score from untreated 48-hour high-induction RPS2-STAMP compared to Torin-1 treated 48-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 $p = 3.9\ e\text{-}98$, Wilcoxon rank-sum), RiboLace (Q1 $p = 8.4\ e\text{-}87$, Wilcoxon rank-sum) and polysome profiling (Q1 $p = 1.2\ e\text{-}54$, Wilcoxon rank-sum).

Figure 4

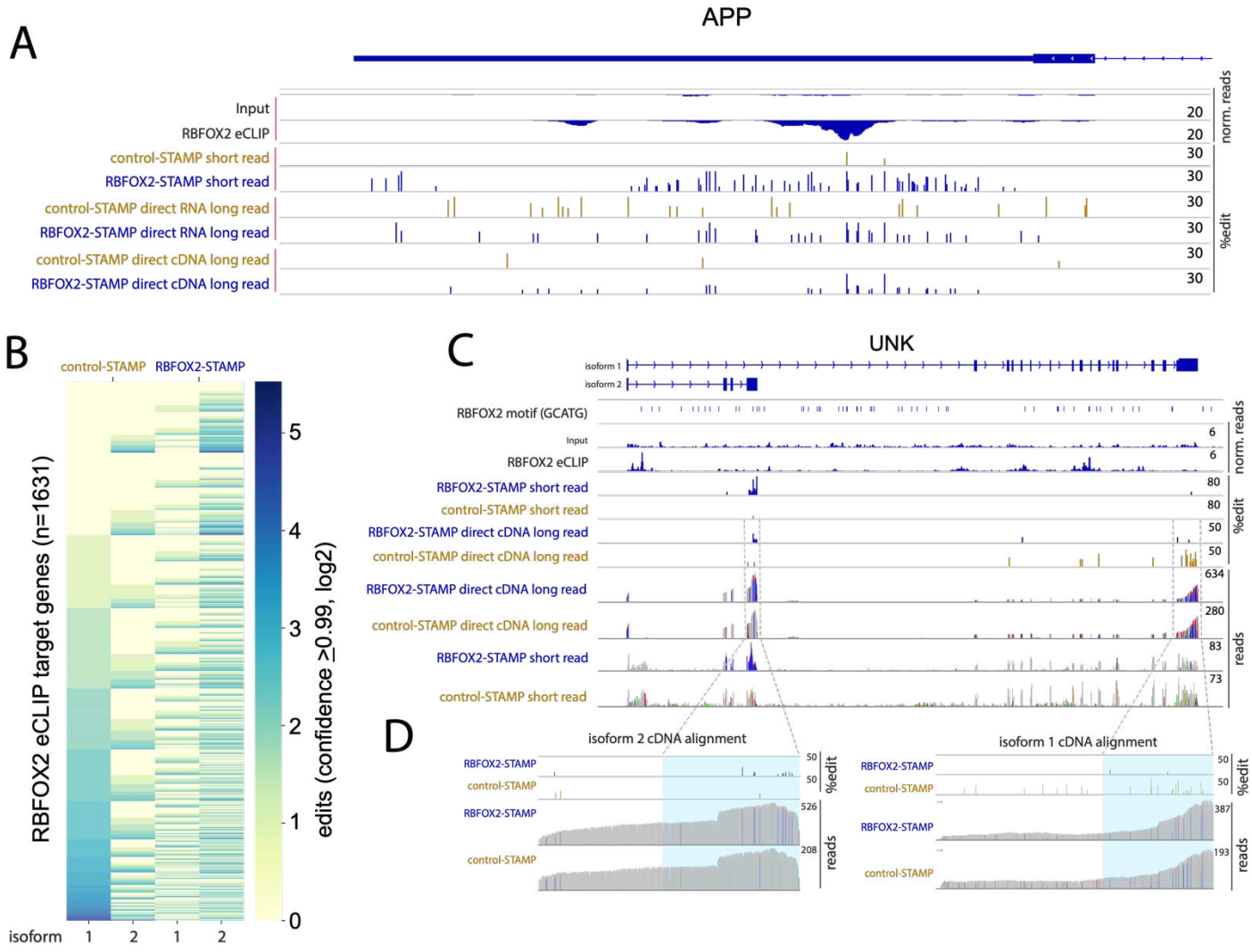

A
APP
Input
RBFOX2 eCLIP
control-STAMP short read
RBFOX2-STAMP short read
control-STAMP direct RNA long read
RBFOX2-STAMP direct RNA long read
control-STAMP direct cDNA long read
RBFOX2-STAMP direct cDNA long read
norm. reads
%edit
B
control-STAMP
RBFOX2-STAMP
RBFOX2 eCLIP target genes (n=1631)
edits (confidence ≥0.99, log2)
isoform
C
UNK
isoform 1
isoform 2
RBFOX2 motif (GCATG)
reads
D
isoform 2 cDNA alignment
isoform 1 cDNA alignment

Figure 4: 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) direct RNA, direct cDNA, and short read (NGS) outputs **B**) Heatmap of control- and RBFOX2-STAMP ≥ 0.99 confidence level edit counts (log2) on the 2 primary isoforms for 1631 RBFOX2 eCLIP target genes ranked by signal form control-STAMP on isoform 1 **C**) IGV tracks showing RBFOX2 eCLIP peaks on the 2 primary isoforms of the target gene *UNK*, compared to control- and RBFOX2-STAMP edit fraction and variant colored read coverage for both long-read direct cDNA and short-read outputs **D**) IGV tracks as in C showing control- and RBFOX2-STAMP direct cDNA outputs in the *UNK* gene focusing only on cDNA alignments for the 2 primary isoforms.

Figure 5

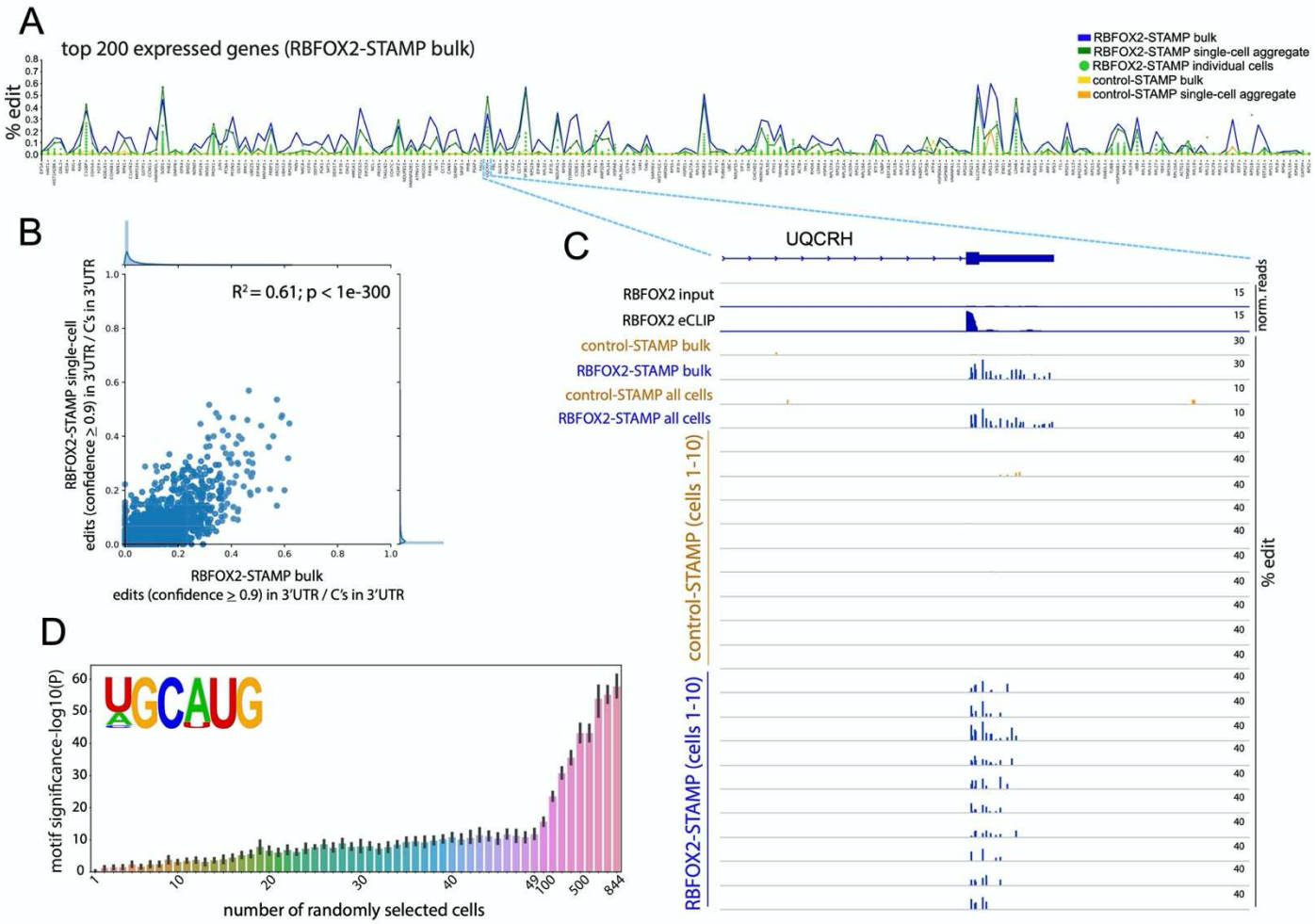

A
top 200 expressed genes (RBFOX2-STAMP bulk)
% edit
RBFOX2-STAMP bulk
RBFOX2-STAMP single-cell aggregate
RBFOX2-STAMP individual cells
control-STAMP bulk
control-STAMP single-cell aggregate
B
R² = 0.61; p < 1e-300
RBFOX2-STAMP single-cell
edits (confidence ≥ 0.9) in 3'UTR / C's in 3'UTR
RBFOX2-STAMP bulk
edits (confidence ≥ 0.9) in 3'UTR / C's in 3'UTR
C
UQCRH
RBFOX2 input
RBFOX2 eCLIP
control-STAMP bulk
RBFOX2-STAMP bulk
control-STAMP all cells
RBFOX2-STAMP all cells
norm. reads
control-STAMP (cells 1-10)
RBFOX2-STAMP (cells 1-10)
% edit
D
UGCAUG
motif significance -log10(P)
number of randomly selected cells

Figure 5: Detection of RBP-RNA targets 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**) Scatterplot of ≥ 0.9 confidence edits divided by all C's across gene 3'UTRs (n = 15832) for bulk and single-cell RBFOX2-STAMP showing high correlation (p value assessed with two-sided test with R beta distribution) **C**) 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 ε scores **D**) -log10 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 6

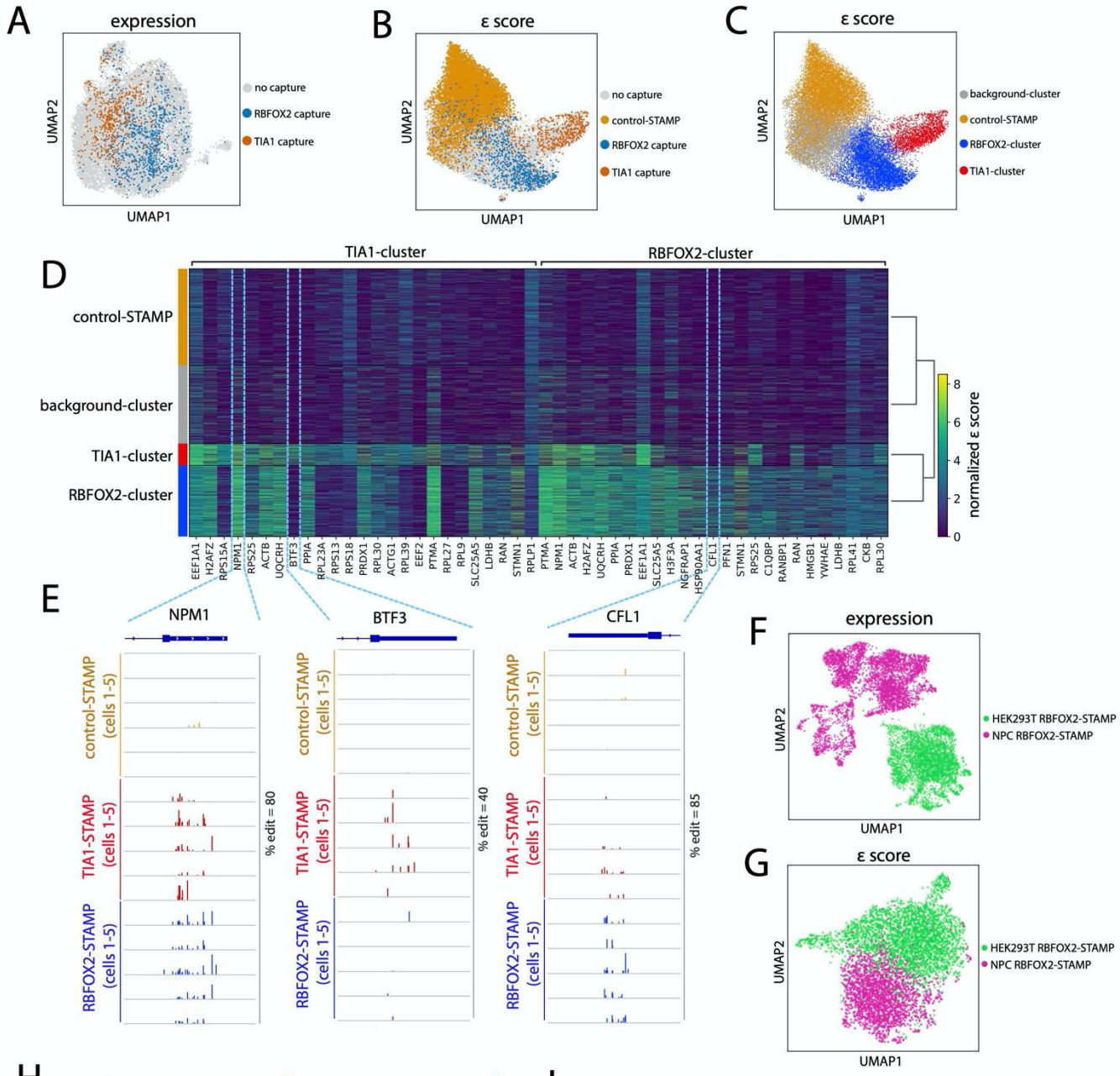

A
expression
UMAP2
UMAP1
no capture
RBFOX2 capture
TIA1 capture
B
ε score
UMAP2
UMAP1
no capture
control-STAMP
RBFOX2 capture
TIA1 capture
C
ε score
UMAP2
UMAP1
background-cluster
control-STAMP
RBFOX2-cluster
TIA1-cluster
D
TIA1-cluster
RBFOX2-cluster
control-STAMP
background-cluster
TIA1-cluster
RBFOX2-cluster
normalized ε score
EEF1A1 H2AFZ RPS15A NPM1 RPS25 ACTB UQCRH BTF3 PPIA RPL23A RPS13 RPS18 PRDX1 RPL30 ACTG1 RPL39 EEF2 PTMA RPL27 RPL9 SLC25A5 LDHB RAN STMN1 RPLP1
PTMA NPM1 ACTB H2AFZ UQCRH PPIA PRDX1 EEF1A1 SLC25A5 H3F3A NGFRAP1 HSP90AA1 CFL1 PFN1 STMN1 RPS25 C1QBP RANBP1 RAN HMGB1 YWHAE LDHB RPL41 CKB RPL30
E
NPM1
BTF3
CFL1
control-STAMP (cells 1-5)
TIA1-STAMP (cells 1-5)
RBFOX2-STAMP (cells 1-5)
% edit = 80
% edit = 40
% edit = 85
F
expression
UMAP2
UMAP1
HEK293T RBFOX2-STAMP
NPC RBFOX2-STAMP
G
ε score
UMAP2
UMAP1
HEK293T RBFOX2-STAMP
NPC RBFOX2-STAMP
H
differentially edited in HEK293T RBFOX2-STAMP
differentially edited in NPC RBFOX2-STAMP
normalized ε score
RPL32 RPL27A RNASEK RPL14 SF3B14 TMSB4X RPS27 RPL28 RPL10 RPL13A
I
differentially edited in HEK293T RBFOX2-STAMP
differentially edited in NPC RBFOX2-STAMP
normalized expression
RPL32 RPL27A RNASEK RPL14 SF3B14 TMSB4X RPS27 RPL28 RPL10 RPL13A
J
RPL14
RPL13A
HEK293T control-STAMP (cells 1-5)
HEK293T RBFOX2-STAMP (cells 1-5)
NPC control-STAMP (cells 1-5)
NPC RBFOX2-STAMP (cells 1-5)
% edit
reads

Figure 6: 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 RBFOX2:TIA1-STAMP cells with capture sequence RBFOX2-STAMP (blue, n = 854) 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-cluster (blue), TIA1-cluster (red) and background-cluster (gray) populations with control-STAMP cells (orange) overlaid **D**) Heatmap of normalized ε score signatures for RBFOX2- and TIA1-cluster cells compared to control-STAMP and background-cluster 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 7

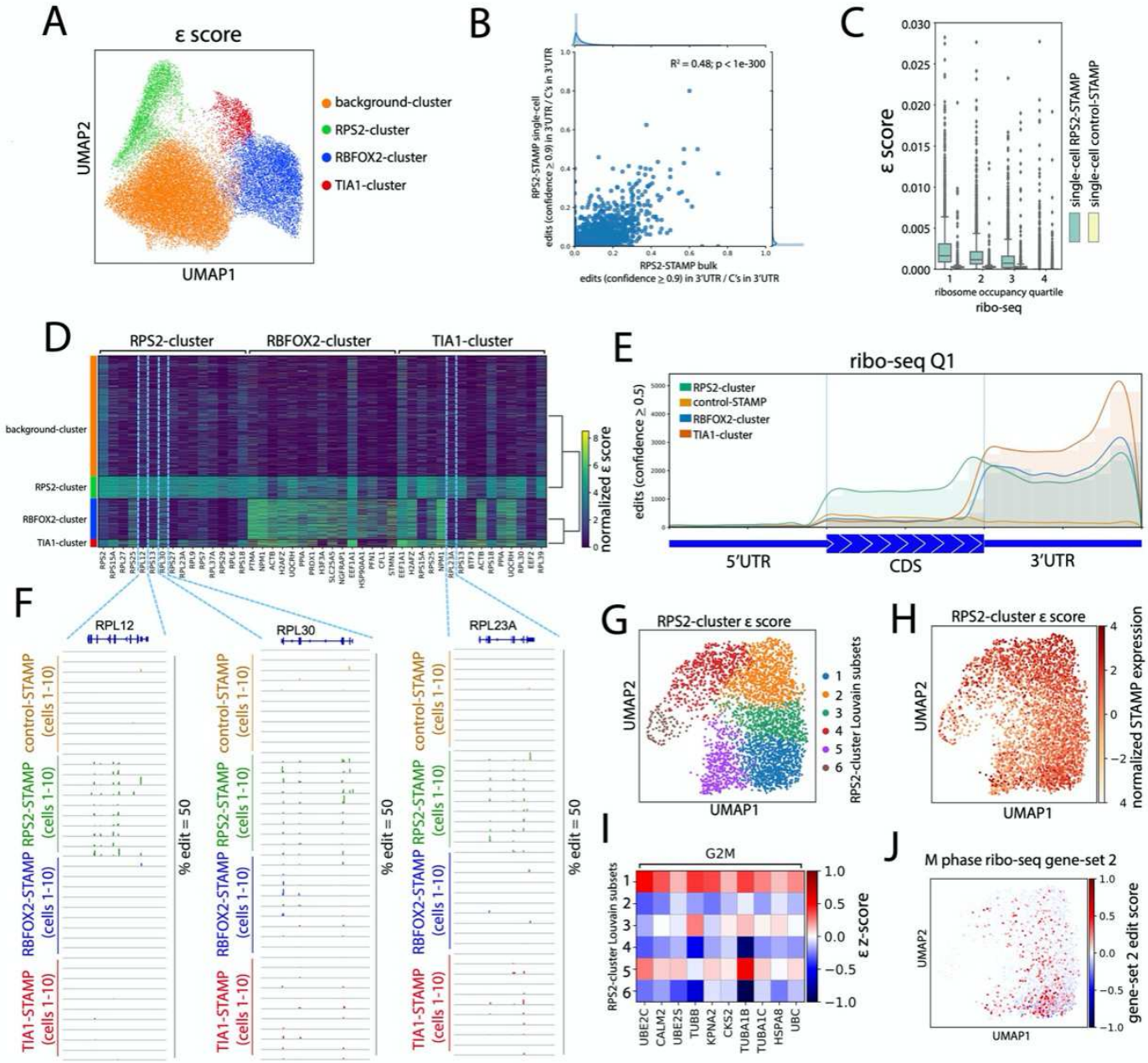

A
ε score
UMAP2
UMAP1
background-cluster
RPS2-cluster
RBFOX2-cluster
TIA1-cluster
B
R² = 0.48; p < 1e-300
RPS2-STAMP single-cell
edits (confidence ≥ 0.9) in 3'UTR / C's in 3'UTR
RPS2-STAMP bulk
edits (confidence ≥ 0.9) in 3'UTR / C's in 3'UTR
C
ε score
single-cell RPS2-STAMP
single-cell control-STAMP
ribosome occupancy quartile
ribo-seq
D
RPS2-cluster
RBFOX2-cluster
TIA1-cluster
background-cluster
RPS2-cluster
RBFOX2-cluster
TIA1-cluster
normalized ε score
E
ribo-seq Q1
RPS2-cluster
control-STAMP
RBFOX2-cluster
TIA1-cluster
edits (confidence ≥ 0.5)
5'UTR
CDS
3'UTR
F
RPL12
RPL30
RPL23A
control-STAMP (cells 1-10)
RPS2-STAMP (cells 1-10)
RBFOX2-STAMP (cells 1-10)
TIA1-STAMP (cells 1-10)
% edit = 50
G
RPS2-cluster ε score
UMAP2
UMAP1
RPS2-cluster Louvain subsets
H
RPS2-cluster ε score
UMAP2
UMAP1
normalized STAMP expression
I
G2M
RPS2-cluster Louvain subsets
ε z-score
UBE2C
CALM2
UBE2S
TUBB
KPNA2
CKS2
TUBA1B
TUBA1C
HSPA8
UBC
J
M phase ribo-seq gene-set 2
UMAP2
UMAP1
gene-set 2 edit score

Figure 7. Ribosome-STAMP reveals ribosome occupancy from individual cells. A) UMAP plot color-coded by ε score Louvain clustering into background-cluster (orange), RBFOX2-cluster (blue), TIA1-cluster (red), and RPS2-cluster (green) from merged 72-hour high-induction STAMP experiments **B**) Scatterplot of ≥ 0.9 confidence edits divided by all C's across gene 3'UTRs for bulk RPS2-STAMP and single-cell RPS2-STAMP (RPS2-cluster, n = 15044 genes) showing high correlation (p value assessed with two-sided test with R beta distribution) **C**) Comparison of gene quartiles ranked by ribosome occupancy (ribo-seq) with ε score from single-cell RPS2-STAMP and single-cell control-STAMP (Q1, Q2 $p < 1e\text{-}300$, Q3 $p = 4.7e\text{-}306$, Wilcoxon rank-sum) **D**) Heatmap of normalized ε score signatures for RPS2-cluster, RBFOX2-cluster and TIA1-cluster cells compared to background-cluster cells on the top 15 differentially edited gene targets **E**) 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 (ribo-seq, n = 4,931 genes) **F**) 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 **G**) UMAP plot of ε score Louvain clustering into 6 RPS2-cluster subsets **H**) UMAP plot as in G showing showing RPS2-STAMP expression **I**) Heatmap of ε score signatures for RPS2-cluster subsets on the top differentially edited gene targets in the G2M cell cycle phase **J**) UMAP analysis of RPS2-cluster subsets showing edit signature for transcripts with enriched M phase ribo-seq signal (Park et al. gene-set 2) edited by RPS2-cluster subset 1 and 5

Figure S1

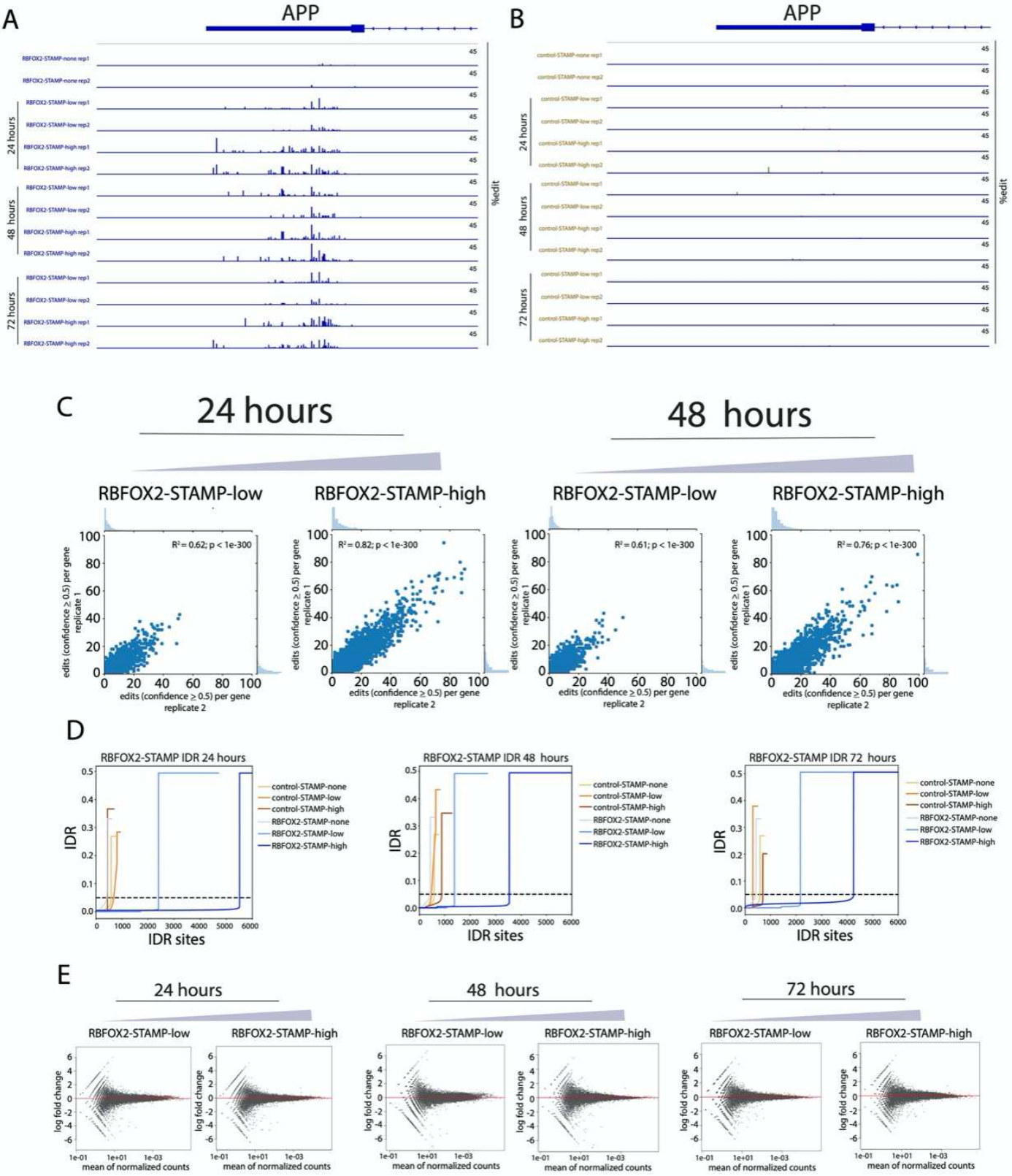

A
APP
RBFOX2-STAMP-none rep1
RBFOX2-STAMP-none rep2
RBFOX2-STAMP-low rep1
RBFOX2-STAMP-low rep2
RBFOX2-STAMP-high rep1
RBFOX2-STAMP-high rep2
24 hours
48 hours
72 hours
%edit
B
APP
control-STAMP-none rep1
control-STAMP-none rep2
control-STAMP-low rep1
control-STAMP-low rep2
control-STAMP-high rep1
control-STAMP-high rep2
24 hours
48 hours
72 hours
%edit
C
24 hours
48 hours
RBFOX2-STAMP-low
RBFOX2-STAMP-high
RBFOX2-STAMP-low
RBFOX2-STAMP-high
R² = 0.62; p < 1e-300
R² = 0.82; p < 1e-300
R² = 0.61; p < 1e-300
R² = 0.76; p < 1e-300
edits (confidence ≥ 0.5) per gene replicate 1
edits (confidence ≥ 0.5) per gene replicate 2
D
RBFOX2-STAMP IDR 24 hours
RBFOX2-STAMP IDR 48 hours
RBFOX2-STAMP IDR 72 hours
IDR
IDR sites
control-STAMP-none
control-STAMP-low
control-STAMP-high
RBFOX2-STAMP-none
RBFOX2-STAMP-low
RBFOX2-STAMP-high
E
24 hours
48 hours
72 hours
RBFOX2-STAMP-low
RBFOX2-STAMP-high
log fold change
mean of normalized counts

Figure S1: Reproducibility of RBFOX2-STAMP, related to Figure 1. A) IGV browser tracks showing RBFOX2-STAMP edit fraction for increasing induction levels and induction times of RBFOX2-STAMP in replicate on the *APP* target gene 3'UTR. **B**) IGV browser tracks as in A showing control-STAMP edit fraction for increasing induction levels and induction times of control-STAMP in replicate on the *APP* target gene 3'UTR. **C**) Scatterplots of ≥ 0.5 confidence level edits per target gene from replicate RBFOX2-STAMP experiments conducted for 24 and 48 hours showing high, dose-dependent correlation (p values assessed with two-sided test with R beta distribution) **D**) Irreproducible Discovery Rate (IDR) analysis comparing ≥ 0.5 confidence edit clusters for increasing levels of RBFOX2-STAMP at 24, 48 and 72 hours. **E**) Differential expression (DEseq2) analysis of RBFOX2-STAMP for increasing levels of RBFOX2-STAMP at 24, 48 and 72 hours

Figure S2

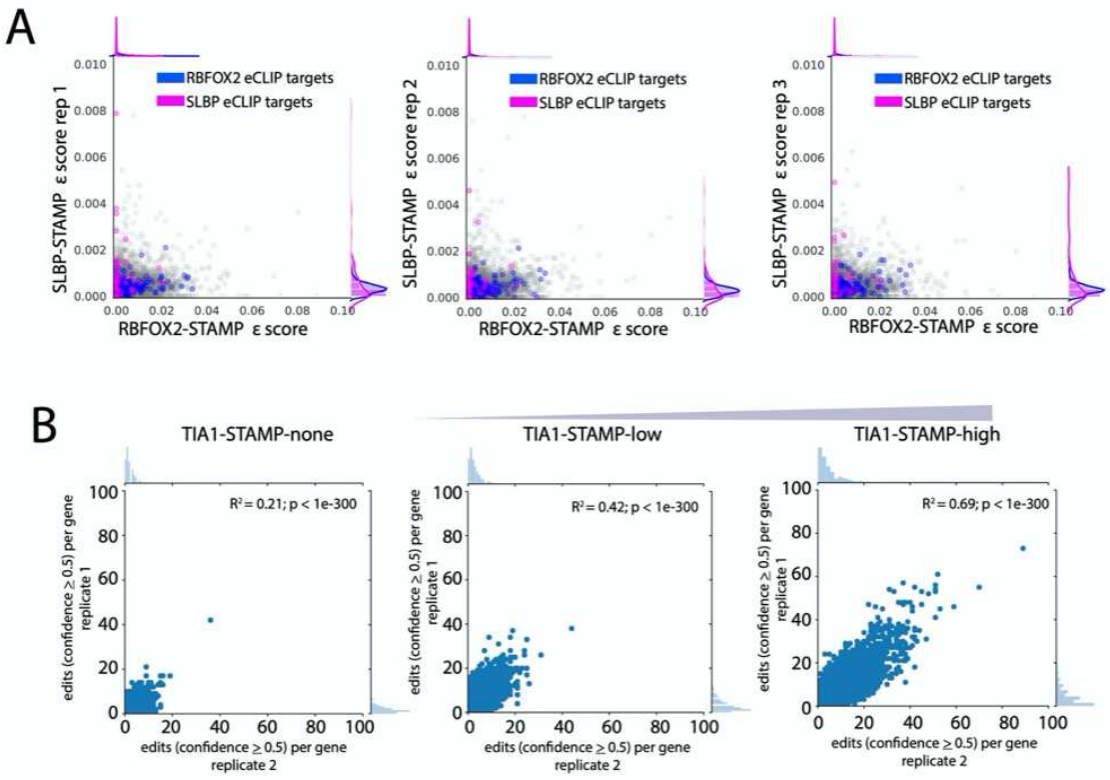
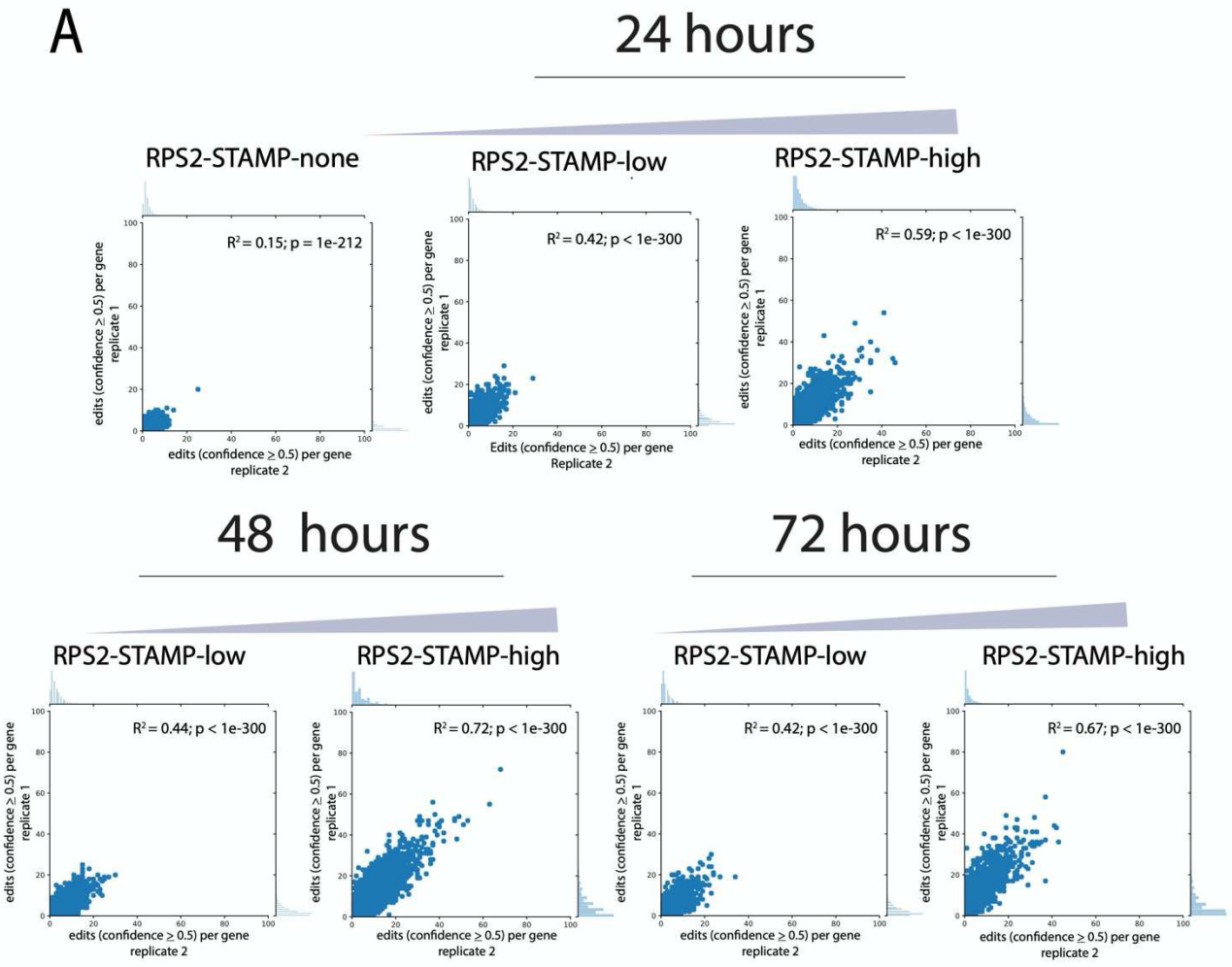

A
RBFOX2 eCLIP targets
SLBP eCLIP targets
SLBP-STAMP ε score rep 1
SLBP-STAMP ε score rep 2
SLBP-STAMP ε score rep 3
RBFOX2-STAMP ε score
B
TIA1-STAMP-none
TIA1-STAMP-low
TIA1-STAMP-high
R² = 0.21; p < 1e-300
R² = 0.42; p < 1e-300
R² = 0.69; p < 1e-300
edits (confidence ≥ 0.5) per gene replicate 1
edits (confidence ≥ 0.5) per gene replicate 2
C
TIA1-STAMP IDR 72 hours
TIA1-STAMP-none
TIA1-STAMP-low
TIA1-STAMP-high
IDR
IDR sites

Figure S2: Reproducibility of SLBP-STAMP, TIA1-STAMP, related to Figure 2 A) Scatterplot comparing ε score for 3 individual replicate 72-hour high-induction SLBP-STAMP experiments with histone gene targets (magenta) highlighted compared to RBFOX2-STAMP ε score with RBFOX2 eCLIP peak containing genes (blue) highlighted across replicates **B**) Scatterplots of ≥ 0.5 confidence level edits per target gene from TIA1-STAMP replicate experiments showing high, dose-dependent correlation (p values assessed with two-sided test with R beta distribution) **C**) Irreproducible Discovery Rate (IDR) analysis comparing 0.5 ≥ confidence level edit clusters for increasing levels of TIA1-STAMP at 72 hours.

Figure S3


A
24 hours
RPS2-STAMP-none
RPS2-STAMP-low
RPS2-STAMP-high
R² = 0.15; p = 1e-212
R² = 0.42; p < 1e-300
R² = 0.59; p < 1e-300
edits (confidence ≥ 0.5) per gene replicate 1
edits (confidence ≥ 0.5) per gene replicate 2
48 hours
72 hours
RPS2-STAMP-low
RPS2-STAMP-high
RPS2-STAMP-low
RPS2-STAMP-high
R² = 0.44; p < 1e-300
R² = 0.72; p < 1e-300
R² = 0.42; p < 1e-300
R² = 0.67; p < 1e-300

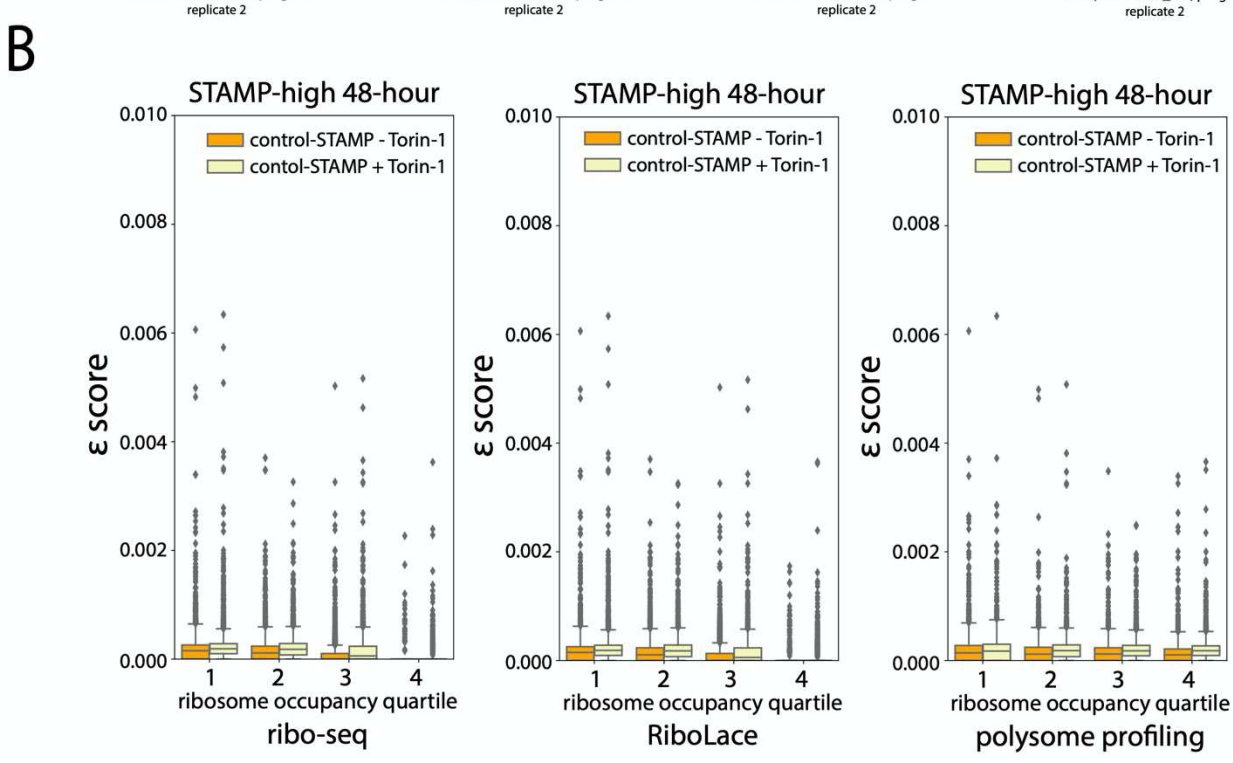

B
STAMP-high 48-hour
control-STAMP - Torin-1
contol-STAMP + Torin-1
control-STAMP + Torin-1
ε score
ribosome occupancy quartile
ribo-seq
RiboLace
polysome profiling

Figure S3: Ribo-STAMP reproducibility and response to mTOR pathway perturbations, related to Figure 3. A) Scatterplots of ≥ 0.5 confidence level edits per target gene from RPS2-STAMP replicate experiments showing high, dose-dependent correlation at 24, 48 and 72 hours (p values assessed with two-sided test with R beta distribution) **B)** Comparison of ε score from untreated 48-hour high-induction control-STAMP compared to Torin-1 treated 48-hour high-induction control-STAMP showing no significant signal reduction for any ribosome occupied quartile genes.

Figure S4

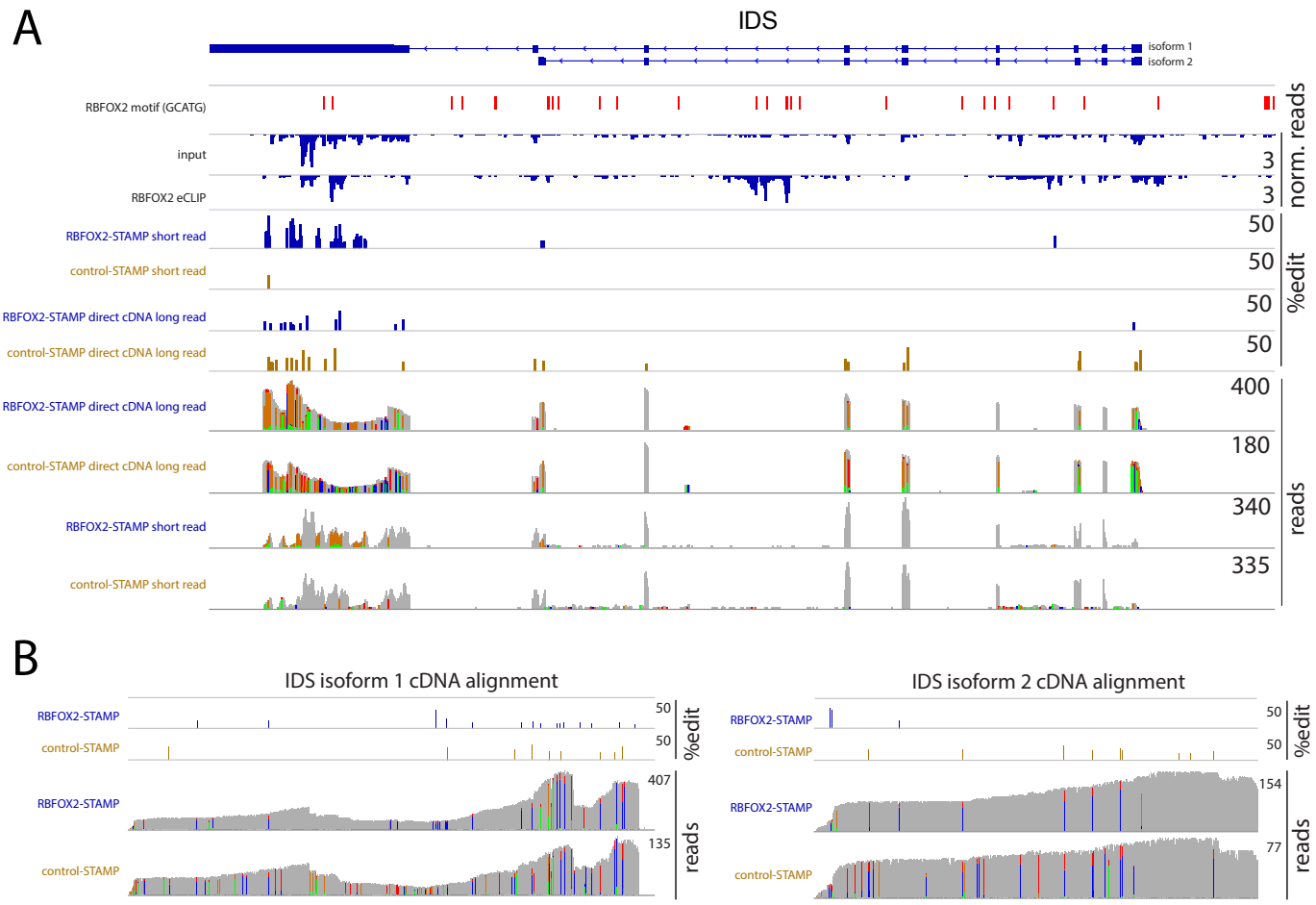

A
IDS
isoform 1
isoform 2
RBFOX2 motif (GCATG)
input
RBFOX2 eCLIP
RBFOX2-STAMP short read
control-STAMP short read
RBFOX2-STAMP direct cDNA long read
control-STAMP direct cDNA long read
RBFOX2-STAMP direct cDNA long read
control-STAMP direct cDNA long read
RBFOX2-STAMP short read
control-STAMP short read
3
3
50
50
50
50
400
180
340
335
norm. reads
%edit
reads
B
IDS isoform 1 cDNA alignment
IDS isoform 2 cDNA alignment
RBFOX2-STAMP
control-STAMP
50
50
407
135
154
77
%edit
reads

Figure S4: Long read sequencing of RBP-STAMP and ribosome-STAMP allows isoform aware detection of binding sites and ribosome occupancy, related to Figure 4: A) IGV tracks showing RBFOX2 eCLIP peaks on the 2 primary isoforms of the target gene *IDS*, compared to 72-hour high-induction control- and RBFOX2-STAMP edit fraction and variant colored read coverage for both long-read direct cDNA and short read (NGS) outputs **B**) IGV tracks as in A showing control- and RBFOX2-STAMP direct cDNA outputs in the *IDS* gene focusing only on cDNA alignments for the 2 primary isoforms.

Figure S5

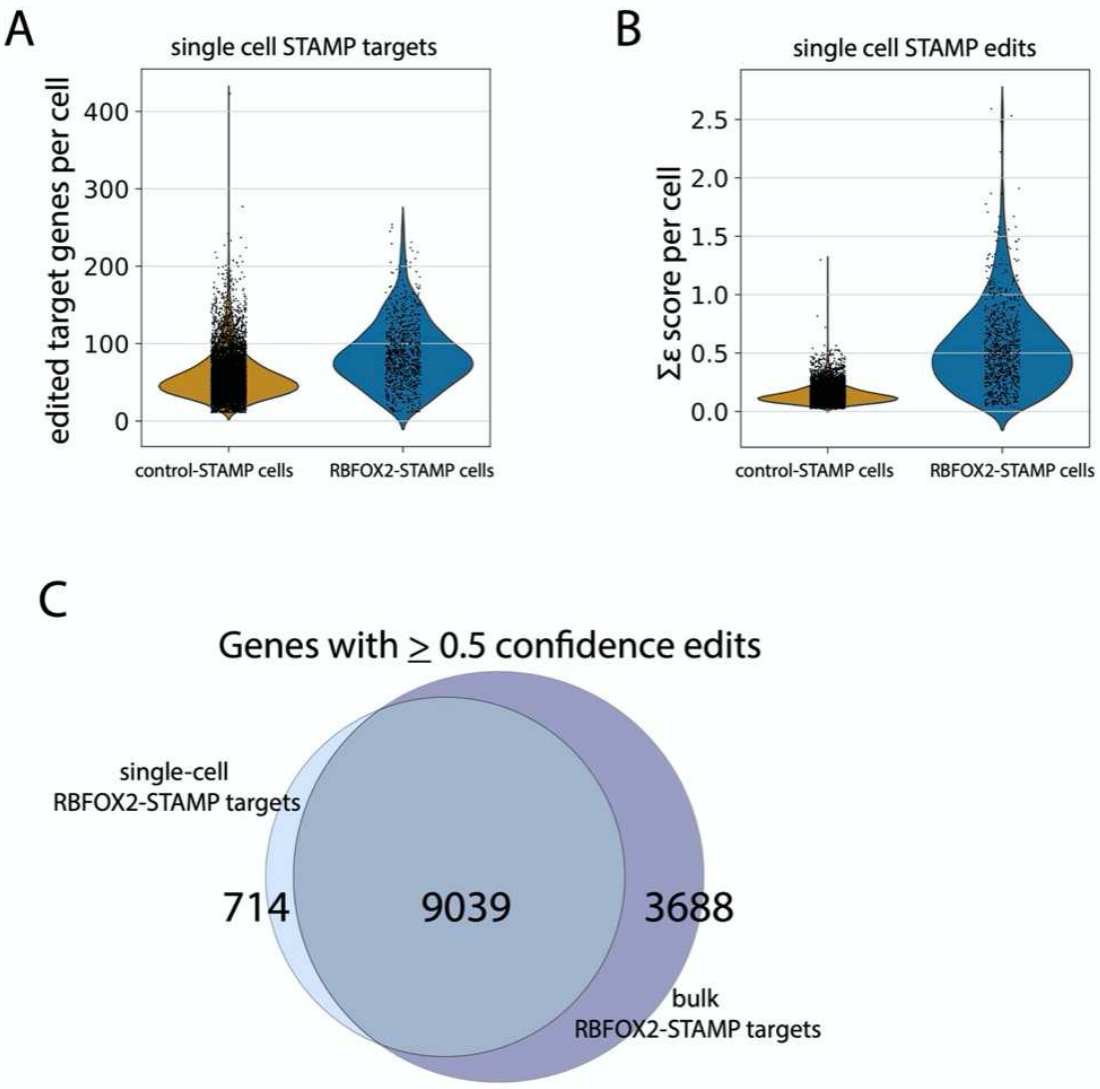

A
single cell STAMP targets
edited target genes per cell
400
300
200
100
0
control-STAMP cells
RBFOX2-STAMP cells
B
single cell STAMP edits
Σε score per cell
2.5
2.0
1.5
1.0
0.5
0.0
control-STAMP cells
RBFOX2-STAMP cells
C
Genes with ≥ 0.5 confidence edits
single-cell
RBFOX2-STAMP targets
714
9039
3688
bulk
RBFOX2-STAMP targets

Figure S5: Comparison of bulk STAMP to single-cell STAMP, related to Figure 5: A) Violin plots summarizing the number of edited genes per cell in 72-hour high-induction control- and RBFOX2-STAMP single-cell experiments **B**) Violin plots summarizing summed ε score per cell for 72-hour high-induction control- and RBFOX2-STAMP single-cell experiments **C**) Venn diagram showing overlap of target genes containing ≥ 0.5 confidence level edits from 72-hour high-induction control- and RBFOX2-STAMP single-cell experiments

Figure S6

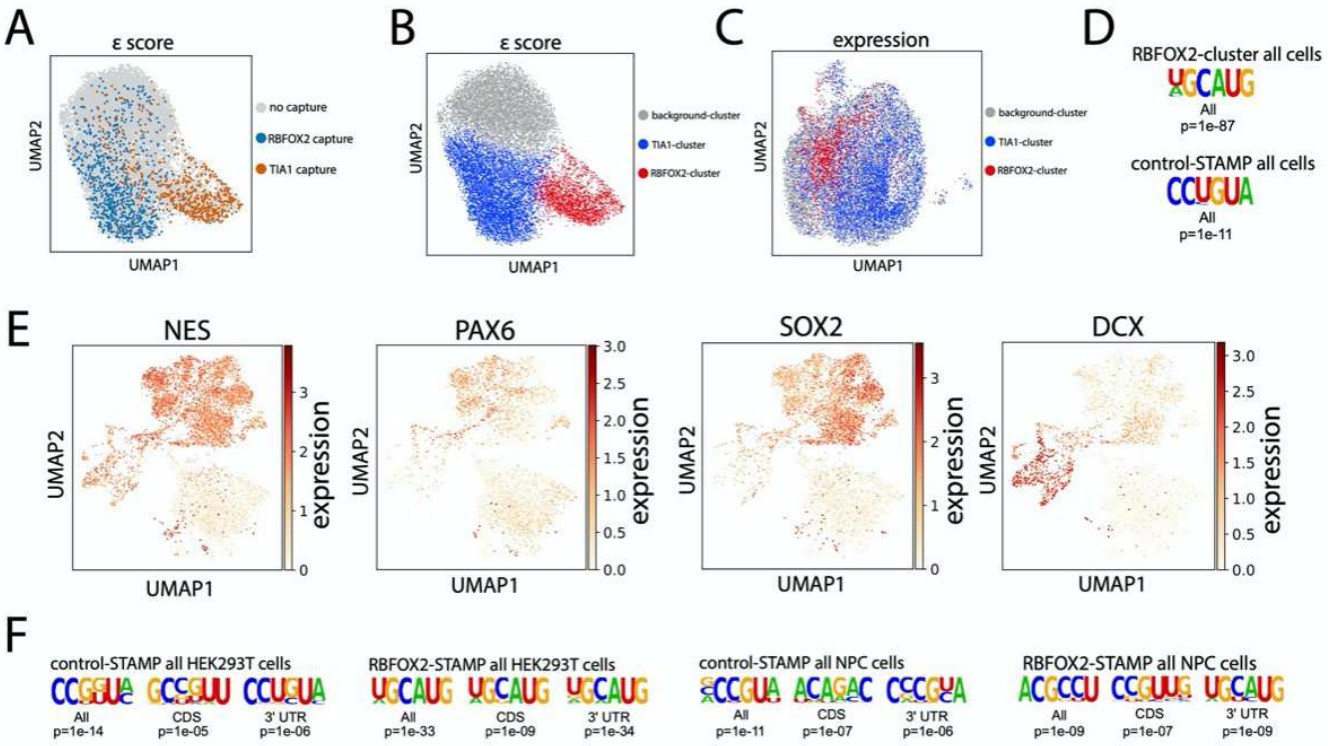

A
ε score
UMAP2
UMAP1
no capture
RBFOX2 capture
TIA1 capture
B
ε score
UMAP2
UMAP1
background-cluster
TIA1-cluster
RBFOX2-cluster
C
expression
UMAP2
UMAP1
background-cluster
TIA1-cluster
RBFOX2-cluster
D
RBFOX2-cluster all cells
All
p=1e-87
control-STAMP all cells
All
p=1e-11
E
NES
PAX6
SOX2
DCX
UMAP2
UMAP1
expression
F
control-STAMP all HEK293T cells
All p=1e-14
CDS p=1e-05
3' UTR p=1e-06
RBFOX2-STAMP all HEK293T cells
All p=1e-33
CDS p=1e-09
3' UTR p=1e-34
control-STAMP all NPC cells
All p=1e-11
CDS p=1e-07
3' UTR p=1e-06
RBFOX2-STAMP all NPC cells
All p=1e-09
CDS p=1e-07
3' UTR p=1e-09

Figure S6 Single-cell RBP-RNA interaction detection by STAMP for multiple RBPs and in multiple cell types, related to figure 6: A) UMAP plot using ε score from RBFOX2-STAMP and TIA1-STAMP mixture with capture sequence RBFOX2-STAMP (blue, n = 854) 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

Figure S7

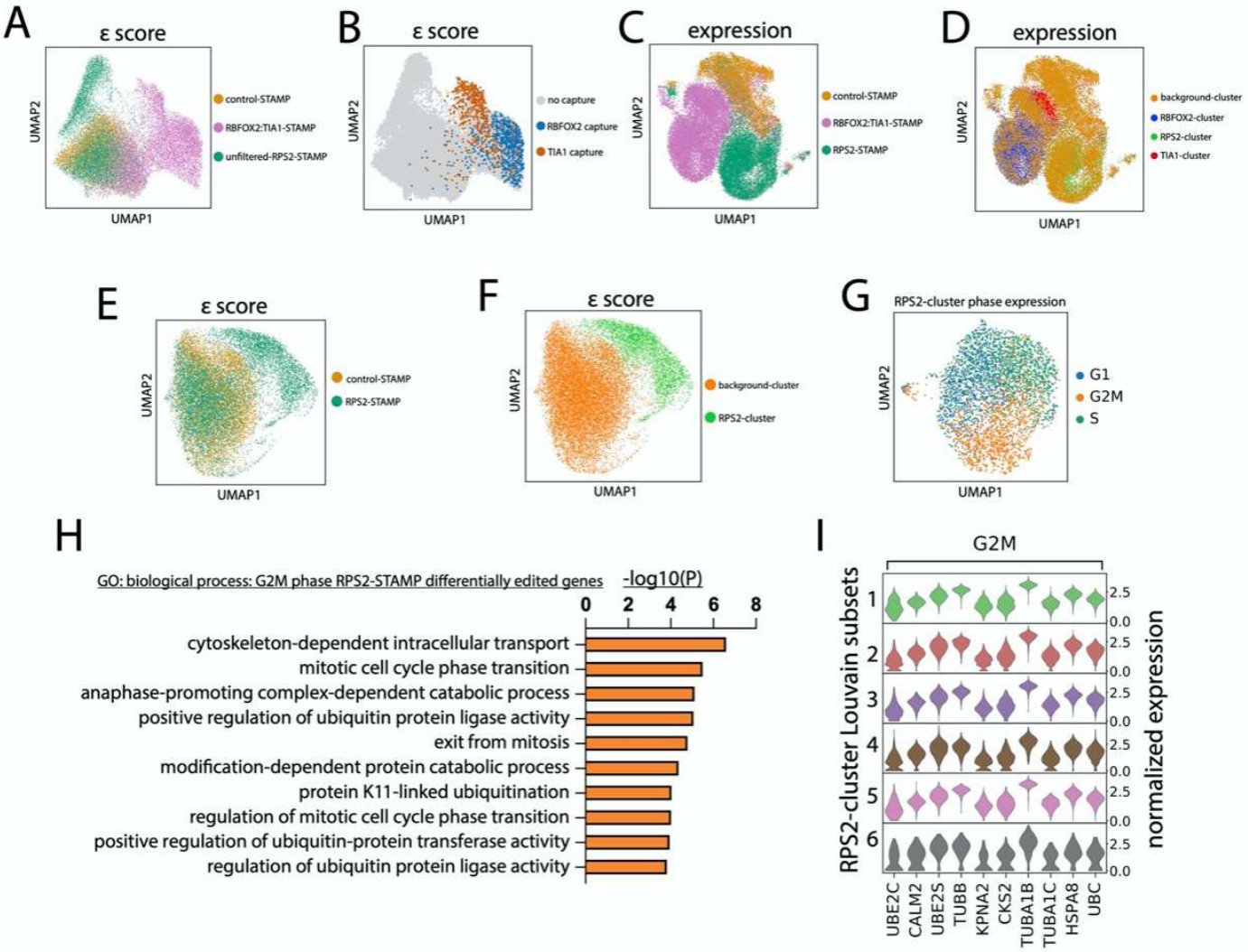

A
ε score
UMAP2
UMAP1
control-STAMP
RBFOX2:TIA1-STAMP
unfiltered-RPS2-STAMP
B
ε score
UMAP2
UMAP1
no capture
RBFOX2 capture
TIA1 capture
C
expression
UMAP2
UMAP1
control-STAMP
RBFOX2:TIA1-STAMP
RPS2-STAMP
D
expression
UMAP2
UMAP1
background-cluster
RBFOX2-cluster
RPS2-cluster
TIA1-cluster
E
ε score
UMAP2
UMAP1
control-STAMP
RPS2-STAMP
F
ε score
UMAP2
UMAP1
background-cluster
RPS2-cluster
G
RPS2-cluster phase expression
UMAP2
UMAP1
G1
G2M
S
H
GO: biological process: G2M phase RPS2-STAMP differentially edited genes
-log10(P)
0
2
4
6
8
cytoskeleton-dependent intracellular transport
mitotic cell cycle phase transition
anaphase-promoting complex-dependent catabolic process
positive regulation of ubiquitin protein ligase activity
exit from mitosis
modification-dependent protein catabolic process
protein K11-linked ubiquitination
regulation of mitotic cell cycle phase transition
positive regulation of ubiquitin-protein transferase activity
regulation of ubiquitin protein ligase activity
I
G2M
RPS2-cluster Louvain subsets
1
2
3
4
5
6
2.5
0.0
normalized expression
UBE2C
CALM2
UBE2S
TUBB
KPNA2
CKS2
TUBA1B
TUBA1C
HSPA8
UBC

Figure S7. Ribosome-STAMP reveals ribosome occupancy from individual cells in different cell cycle phases. Related to Figure 7: A) UMAP analysis of ε score from merged 72-hour high-induction RPS2-STAMP (green), control-STAMP (orange) and RBFOX2:TIA1-STAMP (purple) single-cell experiments **B**) UMAP plot as in A with only capture sequence RBFOX2-STAMP (blue, n = 854) and TIA1-STAMP cells (red, n = 527) highlighted **C**) UMAP analysis of expression from merged 72-hour high-induction RPS2-STAMP (green), control-STAMP (orange) and RBFOX2:TIA1-STAMP (purple) single-cell experiments **D**) UMAP plot of gene expression for ε score Louvain RBFOX2-cluster, TIA1-cluster, RPS2-cluster and background-cluster cell populations **E**) UMAP analysis of ε score for 72-hour high-induction RPS2-STAMP (green), control-STAMP (orange) **F**) UMAP analysis of ε score Louvain clustering into background-cluster and RPS2-cluster subpopulations **G**) UMAP analysis of expression from RPS2-cluster cells, categorizing cells based on cell cycle status (G1, G2M, S) **H**) Gene ontology analysis of the top differentially edited genes from the G2M cell expression RPS2-subgroup **I**) Violin plots summarizing gene expression of the top differentially edited genes in the G2M phase of the cell cycle for the 6 RPS2-cluster Louvain subsets

Table S1. All control-STAMP edit site coordinates and levels; related to Figures 1-4 and Figures S1-S4

Table S2. All RBP-STAMP (RBFOX2, SLBP, TIA1) edit site coordinates and levels; related to Figures 1, 2, and Figures S1, S2

Table S3. Differential gene expression from RBFOX2-STAMP as evaluated by DEseq2

Tabls S4. All Ribo-STAMP (RPS2, RPS3, GSTP1) edit site coordinates and levels; related to Figure 3 and Figure S3

Tabls S5. Differential gene editing between background-cluster and RBFOX2- and TIA1-clusters, related to Figure 6D

Table S6. Differential gene expression between HEK293T cells and NPCs, related to Figure 6F and S6E

Table S7. Differential gene editing between HEK293T cells and NPCs for RBFOX2-STAMP, related to Figure 6G-J

Table S8. Differential gene editing between background-cluster and RPS2-,RBFOX2- and TIA1-clusters, related to Figure 7D

Table S9. Differential gene editing between RPS2-cluster subsets, related to Figure 7G

Table S10. Differential gene editing between cell cycle phase subsets, related to Figure 7I and Figure S7I

Table S11. Oligonucleotides used in this study

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.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

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 SuperSript 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% CO_2. 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% CO_2 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 um 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 1µ/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 2µg expression construct using Fugene HD (Promega) according to manufacterer'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) alongside 1µg/ml doxycycline induction and harvested for RNA after 48 hours 37C incubation.

eCLIP

TIA1 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 or anti-mouse 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).

Generic RNAseq pipeline:

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 (v1.6.3). Genes with zero counts summed across all samples were removed prior to performing differential expression analysis using DESeq2 (v1.26.0) [3].

SAILOR:

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. Sites were transformed into broader "peak" regions by opening a 51-nucleotide window centered on each site.

Edit distribution and e score method details:

Figure 1E: To determine the correlation between edit sites and putative binding, RBFOX2-APOBEC sites were compared to eCLIP binding sites and to the known RBFOX2 motif GCAUG. For RBFOX2 eCLIP peaks containing this motif, distances from the motif to RBFOX2 and APOBEC (background) edit sites were calculated using Bedtools closest (v2.27.1) and plotted as a distribution within a 400bp window. **Figure 1F, G:** To reduce background noise of initial called edit sites, we removed any edit that was also found in our control-STAMP dataset and compared surviving sites on or nearby CDS or 3'UTR 293T eCLIP peaks that contained known motifs (GCAUG for RBFOX2, n=255). For each of these peaks, we opened a window containing the peak location flanked by 25 nucleotides upstream and downstream and merged overlaps. To assess edit enrichment across these windows, we first compared the number of confident edits surrounding each peak against a random background (n trials = 100) of the same length generated using Gencode-defined CDS + 3'UTR regions that 1) belonged to genes that were expressed in the RNASeq dataset, 2) were not within 25 nucleotides of the known motif, 3) were not within 25 nucleotides of a real eCLIP peak (Figure 1F) . We also describe an "ε score" fraction formula: ε 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 e 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. We then used the minimum and maximum ε score to set an evenly spaced range of thresholds by which we used as a filter. Then, we compared ε score around eCLIP peaks to e scores around randomized backgrounds (100 trials). We then calculated the ratio of the fraction of real peaks that passed each threshold to the corresponding fraction of random peaks (Figure 1G). **Figure 1H:** To discover whether or not sites are globally enriched for known binding motifs, we re-calculated the confidence score using the same ε 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).

Figure 2C: Similar analysis was performed as in Figure 1G, using reproducible ENCODE eCLIP peaks derived from K562 cells (ENCSR483NOP). **Figure 2D:** e scores calculated across all exons of genes for both RBFOX2-STAMP and merged SLBP-STAMP triplicates were plotted along the X and Y axes, respectively. Blue highlights genes that were found to be eCLIP (CDS + 3'UTR, n=342) targets in RBFOX2, while purple highlights histone genes containing the UUUU motif. Similar analysis was performed on each SLBP-STAMP triplicate A, B and C (Figure S2A). **Figure 2F:** Similar analysis was performed as in Figure 1G, comparing TIA1-STAMP-high to UUUUU-containing reproducible eCLIP peaks for TIA1 in 293T cells, n=271. **Figure 2G**: Similar to Figure 1H, using TIA1-STAMP-high.

Figure 3C-F: RPS2-STAMP at 0, 50, and 1000ng doxycycline treatments compared to corresponding control-STAMP datasets were compared across genesets taken from GSE112353 and [4]. From GSE112353, 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. e score across exons within these quartiles were compared using a Wilcoxon rank-sum test to determine significance. Similar comparisons were done using normalized occupancy ratios from [4] (using X3 values, which closely approximate native ribosome occupancy levels in 293T cells). **Figure 3G:** 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. Similar to the methods as described in Figure 3C-F, the top quartile transcripts were used, although no expression filtering or transcript-to-gene mapping was needed as transcript-level annotations were required (Q1 n=4,677). **Figure 3H:** Metagene plots were generated in similar fashion to Figure 3G, comparing all replicates of Torin-1 treated RPS2-STAMP, RPS2-STAMP untreated. Since sequencing depth varied among these batches, these datasets were first downsampled to 10 million reads using Samtools view prior to this comparison. **Figure 3I:** Similar to Figure 3C-F, using Torin-1 treated RPS2-STAMP vs RPS2-STAMP untreated. **Figure S3B**: Similar to Figure 3I, using Torin-1 treated control-STAMP, untreated control-STAMP.

Figure 5A-B: To compare single-cell RBFOX2-STAMP with its bulk counterpart, we compared the fraction of edited sites (844 ORF-containing barcodes aggregated and treated as a single “bulk” sample) to the number of C’s found across the 3’UTRs of all expressed genes in the bulk dataset. To do this, we first transformed the list of filtered edits (conf ≥ 0.9) into BAM format with Bedtools bedToBam, then used featureCounts to count the number of edits across 3’UTR regions taken from [cellranger reference data]. We then divided the number of 3’UTR edits by the number of C’s using Bedtools getfasta across the same 3’UTR coordinates. R^2 Pearson correlation across these fractions was computed to determine similarity between the two datasets. **Figure 5D**: 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 as described in Figure 1F-H, 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-calcuated 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). 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 ($-125*\log_2$(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 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 RNA Nanopore Sequencing

500ng of poly-A selected RNA was used as input for the Nanopore direct RNA sequencing kit (SQK-002). RNA was prepared following the manufacturer's protocol. Sequencing was carried out on an Oxford Nanopore Minion-101B using R9.4.1 flow cells for ~48 hours. Data was base called in real time using a Dell Precision 7820 Tower with the Oxford Nanopore Guppy base caller. The RBFOX2-APOBEC sample was sequenced twice and reads pooled. Total reads (in millions) passing default filters were: RBFOX2=1.7, APOBEC_control= 0.8.

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 [5] with default RNA parameters. Theses 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 sites was made by subtracting all sites found in the APOBEC control 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 APOBEC control. 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.

Single cell RNA-seq

For the single cell RNA sequencing of transduced cells. Following 72 hours of doxycycline treatment (1 µg/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 (Aglient Technologies) was performed for QC of the libraries. 3' polyA and feature libraries were sequenced on an Illumina NovaSeq 6000. For 3' polyA de-multiplexing, 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 [6]. 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% mitochrondrial 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 scores for cell cycle genes were computed with the tl.score_genes_cell_cycle 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, e score and motif analysis were run for each cell 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 [6]. 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. e 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 e scores per sample heatmaps were generated with the pl.heatmap and pl.matrixplot functions, respectively. Single cell scores for cell cycle genes and for comparisons with Park et al. [7] were computed with the tl.score_genes_cell_cycle and tl.score_genes function respectively. Differentially edited genes were determined for each set of Louvain (or modified) clusters with the tl.rank_gene_groups function (method='wilcoxon'). Gene ontology enrichment analysis and PPI hub analysis was performed through the Enrichr Gene Ontology enrichment tool (http://amp.pharm.mssm.edu/Enrichr/) [8].

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, E.L., et al., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods, 2016. 13(6): p. 508-14.
3. Love, M.I., W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 2014. 15(12): p. 550.
4. Tan, F.E., et al., A Transcriptome-wide Translational Program Defined by LIN28B Expression Level. Mol Cell, 2019. 73(2): p. 304-313 e3.
5. Li, H., Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 2018. 34(18): p. 3094-3100.
6. Wolf, F.A., P. Angerer, and F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis. Genome Biol, 2018. 19(1): p. 15.
7. Park, J.E., et al., Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. Mol Cell, 2016. 62(3): p. 462-471.
8. Kuleshov, M.V., et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res, 2016. 44(W1): p. W90-7.

Figures

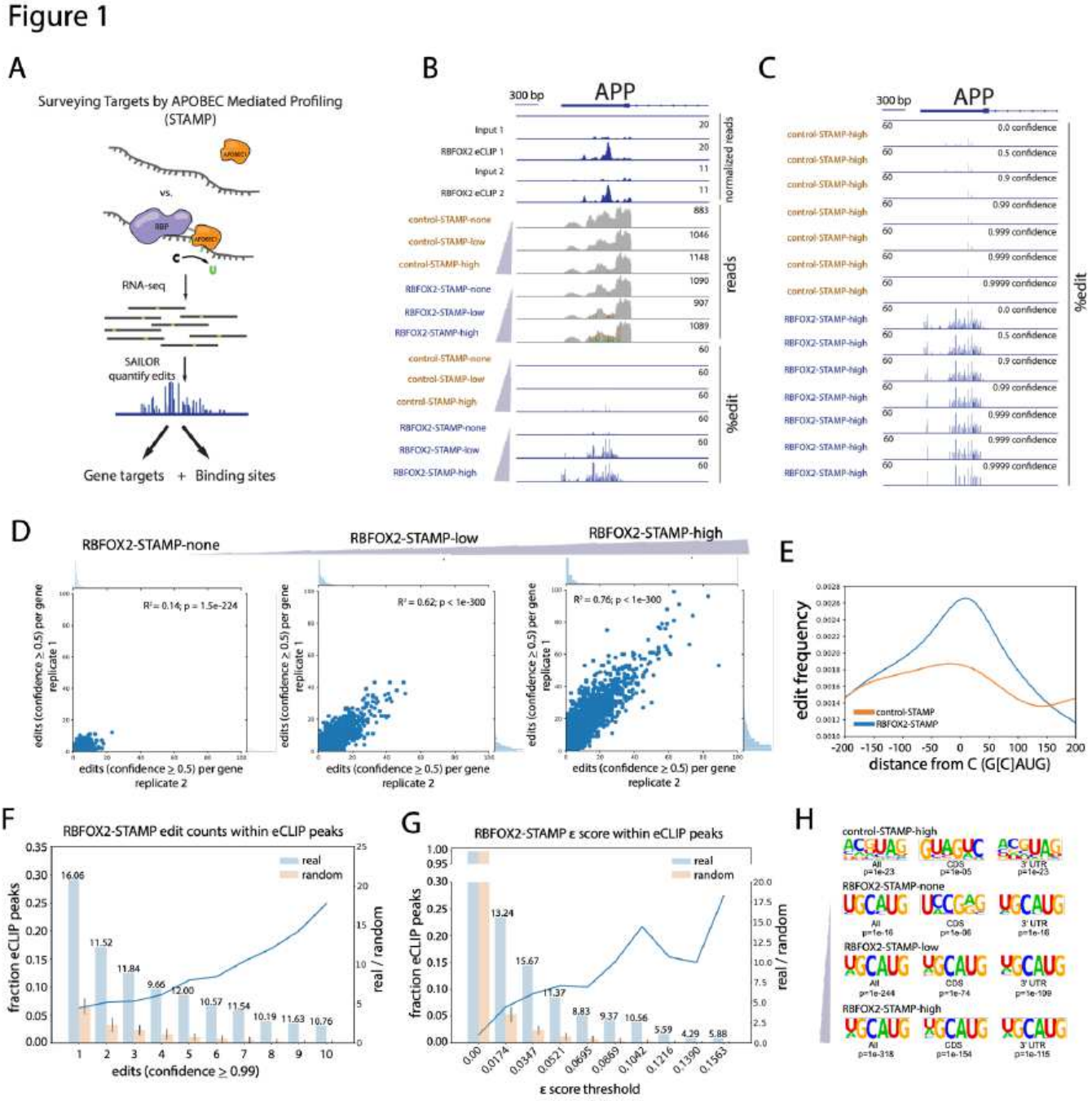


Figure 1

RBFOX2-STAMP enriches C-to-U edits at RBFOX2 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 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 number of edits (confidence level ≥ 0.5) per target (p values assessed with two-sided test with R beta distribution) E) Edit frequency distribution within a 400 bp window flanking RBFOX2 binding-site motifs for RBFOX2-STAMP and control-STAMP (background) F) Fraction of RBFOX2 motif-containing eCLIP peaks in 3'UTR and CDS regions with increasing numbers of ≥ 0.99 confidence edits compared to size-matched randomized regions outside of eCLIP peaks, z-scores reported above bars G) Fraction of RBFOX2 motif-containing eCLIP peaks or random regions, as in G, with increasing edit fraction score (e score) thresholds H) Motif enrichment using HOMER and shuffled background on combined control- or RBFOX2-STAMP ≥ 0.99 confidence level edit windows for increasing RBFOX2-STAMP induction levels.

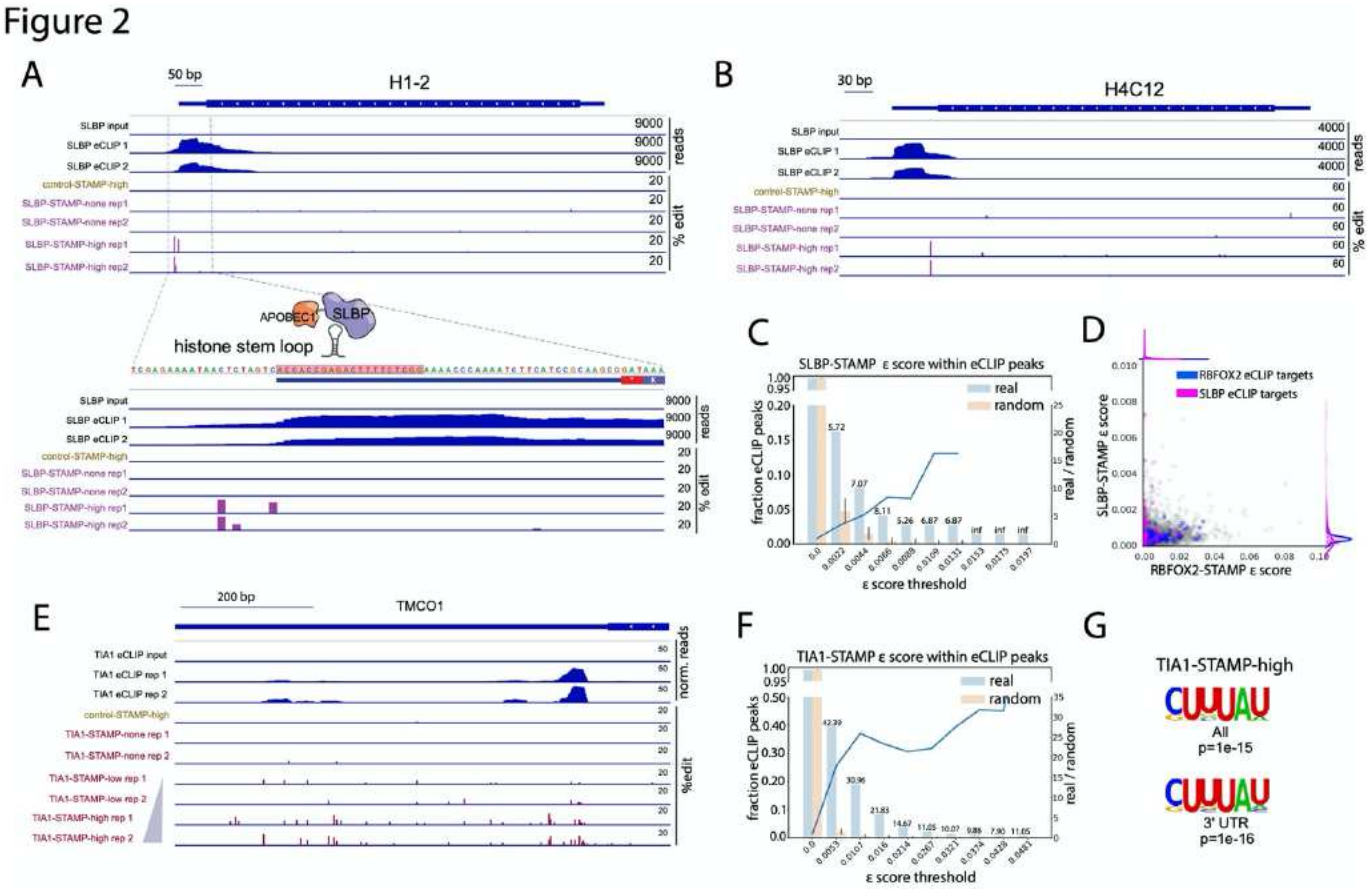


Figure 2

SLBP-STAMP and TIA1-STAMP expression enriches edits on respective targets. A) IGV tracks showing replicate SLBP-STAMP edit fractions at no- and high-induction (doxycycline: 0ng = none or 1µg/ml = high, 72 hours) on the target histone gene H1-2 compared to SLBP eCLIP, highlighting SLBP-bound histone stem loop B) IGV tracks as in A showing replicate SLBP-STAMP edit fractions on the target histone gene H4C12 compared to SLBP eCLIP C) Fraction of SLBP eCLIP peaks or random regions with increasing e

score thresholds, z-scores reported above bars D) Scatterplot comparing e score for merged replicate high-induction SLBP-STAMP with histone gene targets (magenta, n = 89) highlighted compared to high-induction RBFOX2- STAMP e score and RBFOX2 eCLIP peak containing genes (blue, n = 342) highlighted E) browser tracks showing TIA1 eCLIP peaks on the target gene TMCO1, compared with replicate TIA1-STAMP edit fraction for increasing induction levels of TIA1-STAMP F) Fraction of TIA1 eCLIP peaks or random regions with increasing e score thresholds, z-scores reported above bars G) Motif enrichment using HOMER and shuffled background on combined TIA1-STAMP ≥ 0.999 confidence level edit windows for high TIA1-STAMP induction levels

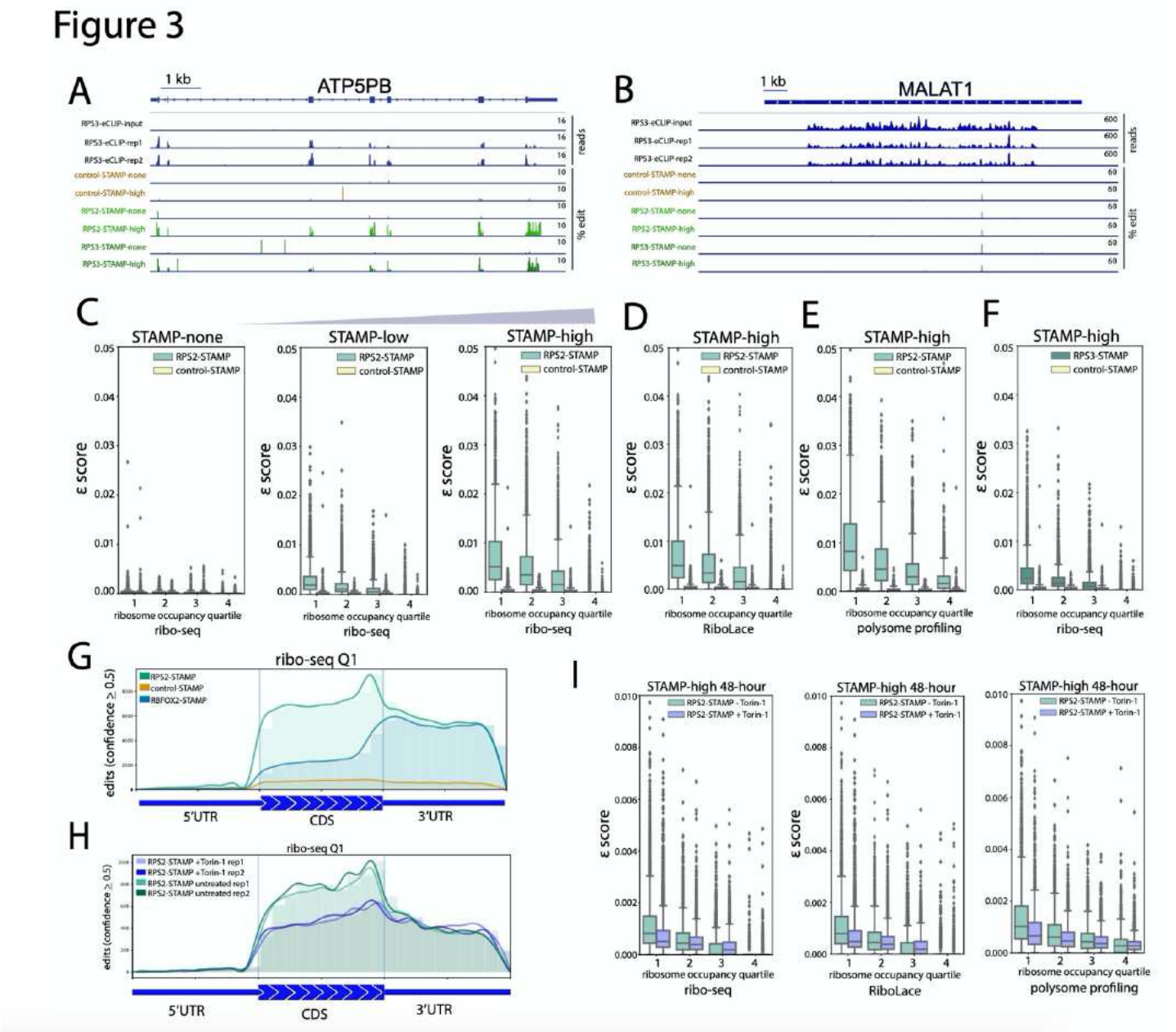


Figure 3

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

highinduction on the ATP5BP 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 or RPS2-, RPS3- STAMP. C) Comparison of gene quartiles ranked by ribosome occupancy (ribo-seq) with e score from increasing levels of RPS2-STAMP and control-STAMP (Q1, Q2 $p < 1e\text{-}300$ for low and high RPS2-STAMP, Wilcoxon rank-sum) D) Comparison of gene quartiles ranked by ribosome occupancy (RiboLace) with e score from high induction of RPS2-STAMP (Q1-Q3 $p < 1e\text{-}300$, Wilcoxon rank-sum) E) Comparison of gene quartiles ranked by ribosome occupancy (polysome profiling) with e score from high induction of RPS2-STAMP (Q1-Q4 $p < 1e\text{-}300$, Wilcoxon rank-sum) F) Comparison of gene quartiles ranked by ribosome occupancy (ribo-seq) with e score from high induction of RPS3-STAMP (Q1,Q2 $p < 1e\text{-}300$, Wilcoxon rank-sum) G) Metagene plot showing edit (≥ 0.5 confidence level) 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) H) Metagene plot showing edit (≥ 0.5 confidence level) distribution for untreated 48-hour high-induction RPS2-STAMP compared to replicate Torin-1 treated 48-hour high-induction RPS2-STAMP across 5'UTR, CDS and 3'UTR gene regions for the top quartile of ribosome occupied genes (ribo-seq, n = 4931 genes) I) Comparison of e score from untreated 48-hour high-induction RPS2-STAMP compared to Torin-1 treated 48- 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 $p = 3.9\ e\text{-}98$, Wilcoxon rank-sum), RiboLace (Q1 $p = 8.4\ e\text{-}87$, Wilcoxon rank-sum) and polysome profiling (Q1 $p = 1.2\ e\text{-}54$, Wilcoxon rank-sum).

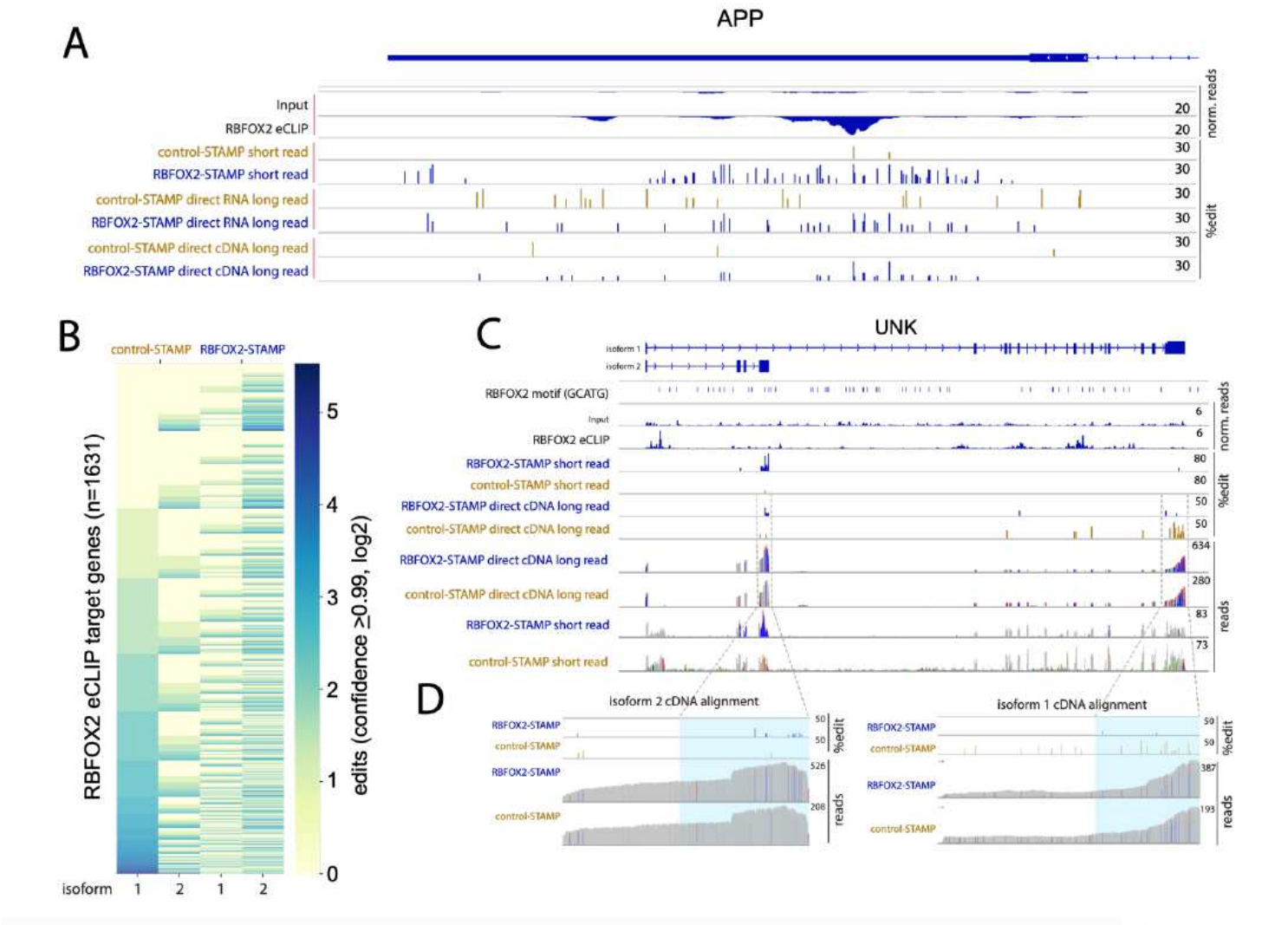


Figure 4

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) direct RNA, direct cDNA, and short read (NGS) outputs B) Heatmap of control- and RBFOX2-STAMP ≥ 0.99 confidence level edit counts (log2) on the 2 primary isoforms for 1631 RBFOX2 eCLIP target genes ranked by signal form control-STAMP on isoform 1 C) IGV tracks showing RBFOX2 eCLIP peaks on the 2 primary isoforms of the target gene UNK, compared to control- and RBFOX2-STAMP edit fraction and variant colored read coverage for both long-read direct cDNA and short-read outputs D) IGV tracks as in C showing control- and RBFOX2-STAMP direct cDNA outputs in the UNK gene focusing only on cDNA alignments for the 2 primary isoforms.

Figure 5

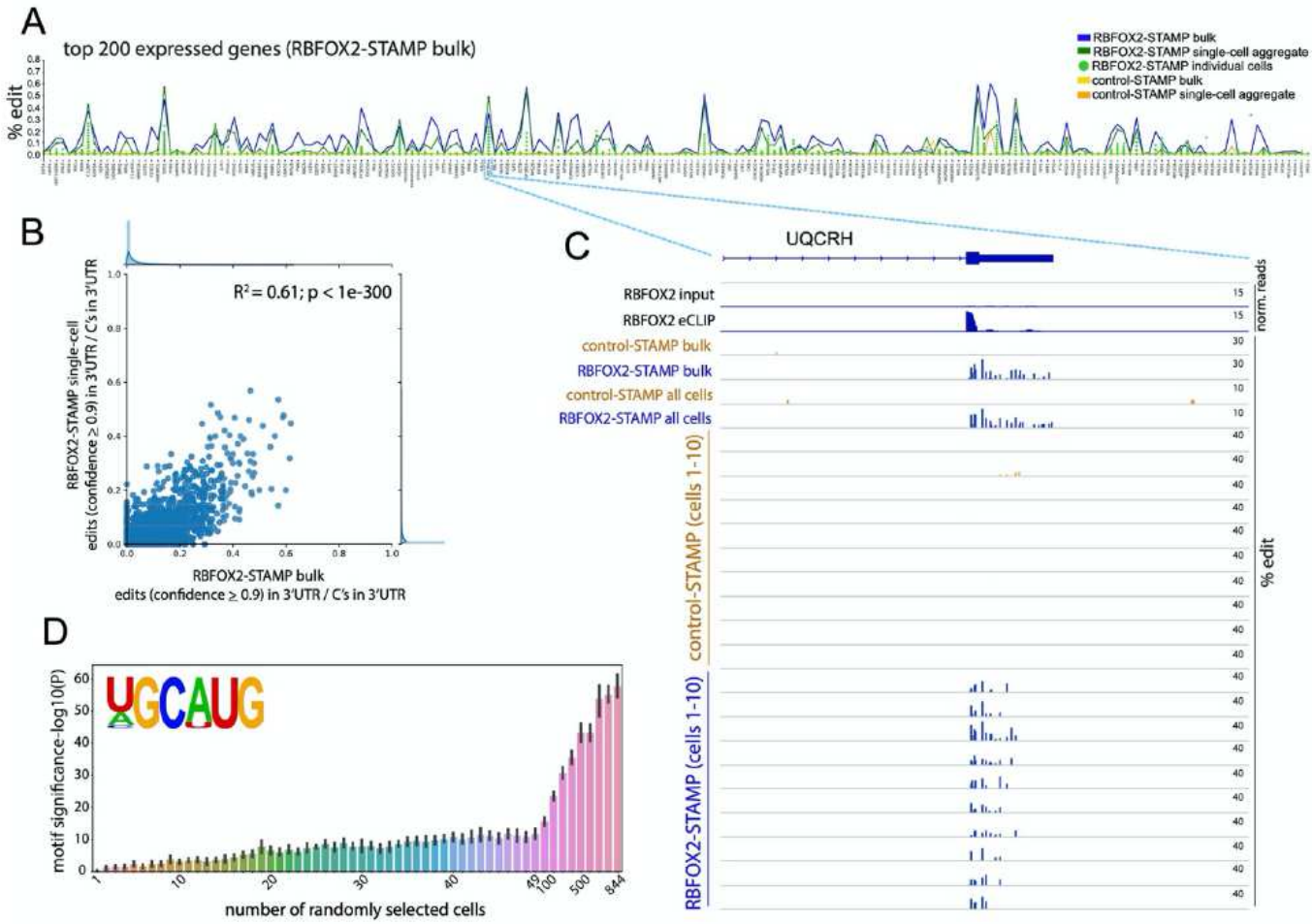


Figure 5

Detection of RBP-RNA targets 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) Scatterplot of ≥ 0.9 confidence edits divided by all C's across gene 3'UTRs (n = 15832) for bulk and single-cell RBFOX2-STAMP showing high correlation (p value assessed with two-sided test with R beta distribution) C) 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 D) -log10 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.

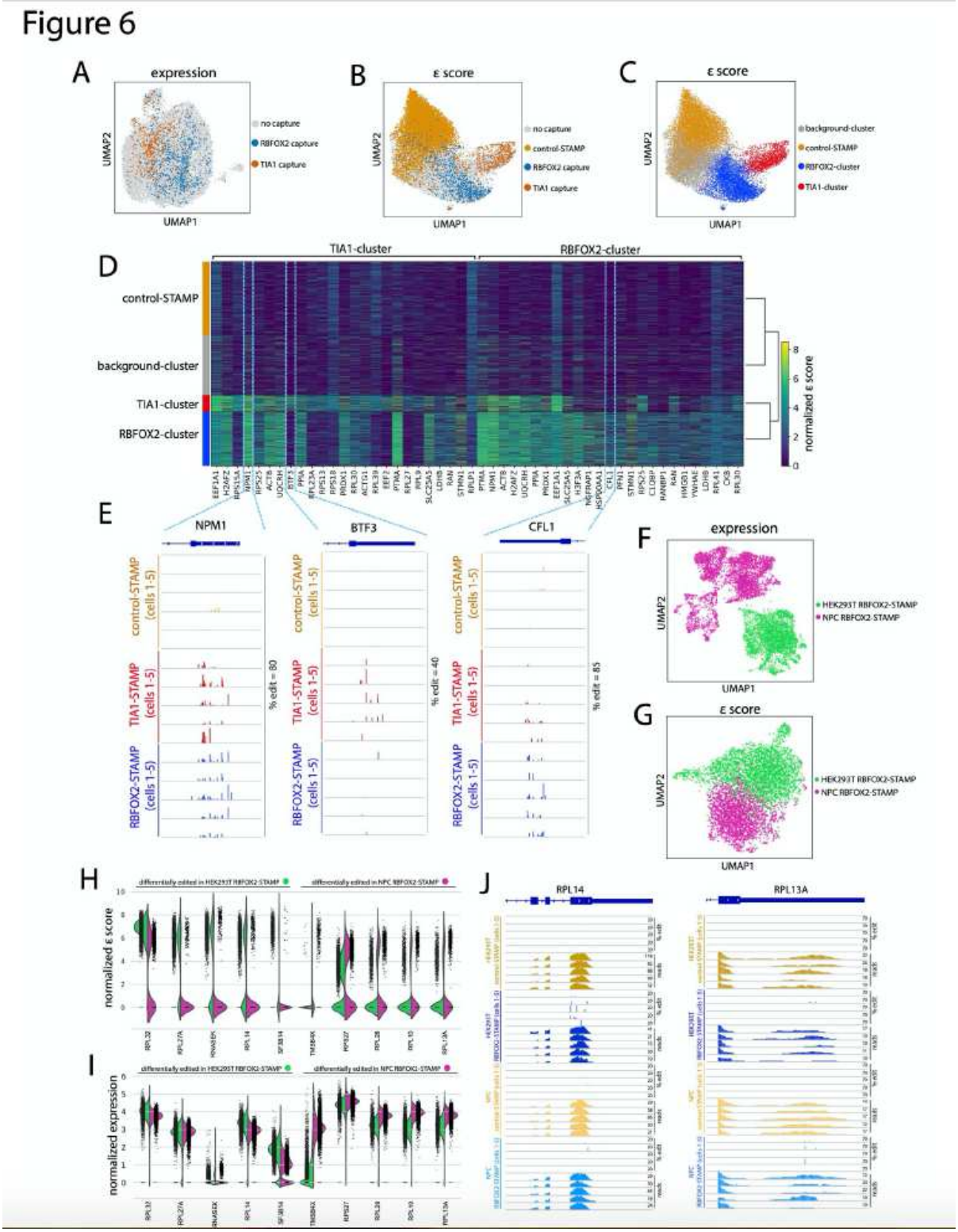


Figure 6

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 RBFOX2:TIA1-STAMP cells with capture sequence RBFOX2-STAMP (blue, n = 854) and TIA1-STAMP cells (red, n = 527) highlighted B) UMAP analysis using e score rather than gene expression after merging 72-hour high-induction control-STAMP cells (orange) C) UMAP plot as in B color-coded by e score Louvain clustering

into RBFOX2-cluster (blue), TIA1-cluster (red) and background-cluster (gray) populations with control-STAMP cells (orange) overlaid D) Heatmap of normalized e score signatures for RBFOX2- and TIA1-cluster cells compared to control-STAMP and background-cluster 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 e 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 e score H) e 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 e score J) IGV browser tracks showing edit fractions and read coverage for the top 5 control- and RBFOX2- STAMP cells (ranked by summed e scores) on the RPL14 and RPL13A gene targets

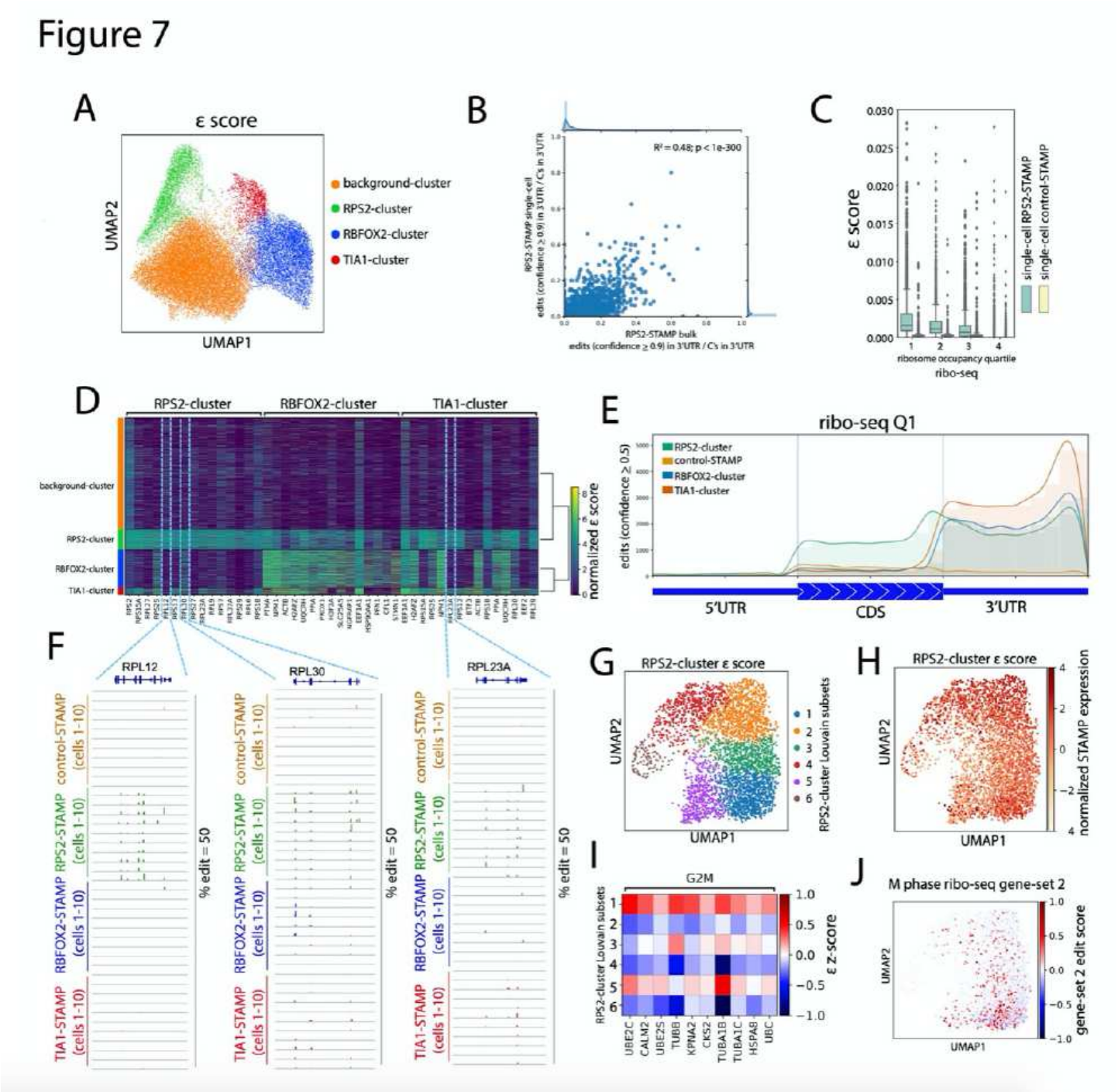


Figure 7

Ribosome-STAMP reveals ribosome occupancy from individual cells. A) UMAP plot color-coded by e score Louvain clustering into background-cluster (orange), RBFOX2-cluster (blue), TIA1-cluster (red), and RPS2-cluster (green) from merged 72-hour high-induction STAMP experiments B) Scatterplot of ≥ 0.9 confidence edits divided by all C's across gene 3'UTRs for bulk RPS2-STAMP and single-cell RPS2-STAMP (RPS2-cluster, n = 15044 genes) showing high correlation (p value assessed with two-sided test with R beta distribution) C) Comparison of gene quartiles ranked by ribosome occupancy (ribo-seq) with e score from single cell RPS2-STAMP and single-cell control-STAMP (Q1, Q2 p < 1e-300, Q3 p = 4.7e-306, Wilcoxon rank-sum) D) Heatmap of normalized e score signatures for RPS2-cluster, RBFOX2-cluster and

TIA1-cluster cells compared3 to background-cluster cells on the top 15 differentially edited gene targets E) 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 (ribo seq, n = 4,931 genes) F) IGV browser tracks showing edit fractions for the top 10 control-, RPS2-, RBFOX2-, and TIA1-STAMP cells (ranked by summed e scores) on the RPL12, RPL30 and RPL23A gene targets G) UMAP plot of e score Louvain clustering into 6 RPS2-cluster subsets H) UMAP plot as in G showing showing RPS2- STAMP expression I) Heatmap of e score signatures for RPS2-cluster subsets on the top differentially edited gene targets in the G2M cell cycle phase J) UMAP analysis of RPS2-cluster subsets showing edit signature for transcripts with enriched M phase ribo-seq signal (Park et al. gene-set 2) edited by RPS2-cluster subset 1 and 5

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx
- TableS9.xlsx
- TableS10.xlsx
- TableS11.xlsx