



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

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

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 ultraviolet cross-linking or immunoprecipitation and, when coupled with single-cell capture, can identify RBP-specific and cell-type-specific RNA-protein interactions for multiple RBPs and cell types in single, pooled experiments. Pairing STAMP with long-read sequencing yields RBP target sites in an isoform-specific manner. Finally, Ribo-STAMP leverages small ribosomal subunits to measure transcriptome-wide ribosome association in single cells. STAMP enables the study of RBP-RNA interactions and translational landscapes with unprecedented cellular resolution.

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 states³. Immunoprecipitation (IP)-based technologies coupled with high-throughput sequencing such as RNA immunoprecipitation (RIP) and cross-linking immunoprecipitation (CLIP) are commonly used to identify RBP targets and binding sites across the transcriptome⁴. While RIP-seq is useful for identifying gene targets of an RBP, CLIP-seq can resolve binding sites within different regions of a given target gene, which lends insight into binding functionality, and allows for the discovery of sequence motifs recognized by specific RBPs^{5,6}. The eukaryotic ribosome is itself composed of a collection of RBPs that can interact directly with mRNA coding sequences⁷. Ribosome profiling methods such as Ribo-seq have become a mainstay in the evaluation of transcriptome-scale ribosome occupancy^{8,9}. Unfortunately, CLIP and ribosome profiling experimental protocols are labor intensive, usually require sizable amounts of input material and transcript fragmentation^{10,11}, prohibiting single-cell and long-read platform applications. While there has been rapid progress in single-cell measurements of chromatin accessibility¹², gene expression^{13,14} and surface protein levels^{15,16}, there is currently no available technology for measuring RBP-mRNA and ribosome-mRNA interactions at single-cell or isoform-aware resolution.

Recent studies have circumvented the need for IP for detecting RBP-RNA interactions by utilizing fusions of RNA-editing or RNA-modifying modules to RBPs of interest to label RNA targets^{17–21}. Internal RNA target labeling has been accomplished by the target of RBPs 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 adenosine to inosine (A-to-I) edits^{17,19,20,22,23}. These ADAR-mediated approaches

have been used to obtain RBP targets from low-input material, but are limited by the sparsity of double-stranded regions proximal to RBP binding sites that are required for ADAR-mediated (A-to-I) editing²⁴.

APOBEC1 is a cytosine deaminase that catalyzes RNA cytosine-to-uracil (C-to-U) conversion on single-stranded RNA substrates²⁵. Recently APOBEC1 was fused to the m6A-binding YTH domain to identify m6A modification sites on RNAs²⁶. This approach, termed deamination adjacent to RNA modification target (DART-seq), identified YTH-domain-recognized m6A modifications on single-stranded mRNAs. However, it was unclear if APOBEC1 fusions would work with general classes of RBPs or even ribosomes. We reasoned that fusion of APOBEC1 to such full-length RBPs could generalize robust, IP-free identification of RBP targets across functional RNA-interaction categories using extremely low or even single-cell input. Here, we demonstrate the power of such an approach for detecting RBP-RNA targets at the single-cell and single-molecule level. We developed an integrated experimental and computational framework termed STAMP, which extends the DART-seq approach to demonstrate the discovery of RBP-RNA sites at isoform-specific and single-cell resolution, the deconvolution of targets for multiplexed RBPs, and the cell type-specific binding of an RBP in a heterogeneous mixture of cell types. Furthermore, by applying STAMP with specific ribosome subunits, we extend this approach for single-cell detection of ribosome association while simultaneously measuring gene expression.

Results

STAMP identifies RBP binding sites without immunoprecipitation. Our strategy for IP-free detection of RBP targets involves fusing full-length RBPs of interest to the cytidine deaminase enzyme APOBEC1, which is known to catalyze C-to-U editing on single-stranded RNA targets (Fig. 1a). Upon expression of an RBP-APOBEC1 fusion protein (RBP-STAMP), RBPs direct the

¹Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. ²Stem Cell Program, University of California San Diego, La Jolla, CA, USA. ³Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. ⁴Pacific Biosciences (PacBio) of California, Menlo Park, CA, USA. ✉e-mail: geneyeo@ucsd.edu

deaminase module to their RNA targets leading to C-to-U base conversion proximal to RBP binding sites. These mutations (edits) are resolved using high-throughput RNA-sequencing (RNA-seq) approaches and quantified using the SAILOR analysis pipeline²⁷, which we modified to identify and assign a confidence value for C-to-U mismatches using a beta distribution that factors both site coverage and editing percentage following removal of annotated single-nucleotide polymorphism (SNPs)²⁸ (Methods).

To determine the utility of the STAMP approach, we fused APOBEC1 to the C terminus of the RBP RBFOX2 (refs. 29–31) 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 (50 ng ml⁻¹ doxycycline) and higher (1 µg ml⁻¹ doxycycline) levels of RBFOX2–STAMP for 72 h had enriched C-to-U edit clusters on the 3′ untranslated region (3′ UTR) of the known RBFOX2 target *APP* mRNA, and these edit clusters coincided with reproducible RBFOX2 binding sites as detected by enhanced CLIP (eCLIP) of either endogenous RBFOX2 (ref. 32) or the RBFOX2–APOBEC1 fusion (Fig. 1b). Uninduced RBFOX2–STAMP, or control–STAMP (APOBEC1 only) at low and high induction, had few or no detectable C-to-U edits in the same region, indicating target specificity. RBFOX2–STAMP-induced edits within this *APP* 3′ UTR target region were tenfold to 25-fold more frequent than background control–STAMP edits at 0.9 and 0.999 (SAILOR) confidence thresholds, respectively (Fig. 1c and Supplementary Tables 1 and 2). 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–STAMP and RBFOX2–STAMP with low and high doxycycline inductions for 24, 48 and 72 h. The number of edited reads (E) on each target gene, normalized to read depth and gene length (edited reads per kilobase of transcripts per million mapped reads (EPKM)), was highly reproducible and correlations between replicates improved substantially upon induction ($R^2=0.32$ at no doxycycline treatment; $R^2=0.72$ and 0.83 at low and high doxycycline, respectively; Fig. 1d). Irreproducible discovery rate (IDR) analysis³³ also revealed reproducible windows with edits for RBFOX2–STAMP, and the number of these reproducible edits also increased with doxycycline induction of RBFOX2–STAMP (Extended Data Fig. 1a). We also evaluated the effects of RBFOX2–STAMP editing on target transcript levels by conducting differential gene expression analysis

on low-induction and high-induction RBFOX2–STAMP at multiple time points and detected negligible changes in cellular gene expression compared to uninduced controls (see Extended Data Fig. 1b for results of the 72-h time point, which is similar to the 24 h and 48 h results; Supplementary Table 3). We observed expected basal leakiness of the doxycycline system, but with induction, RBFOX2–APOBEC1 mRNA levels increased to within 1.5-fold of the endogenous RBFOX2 levels (Fig. 1e).

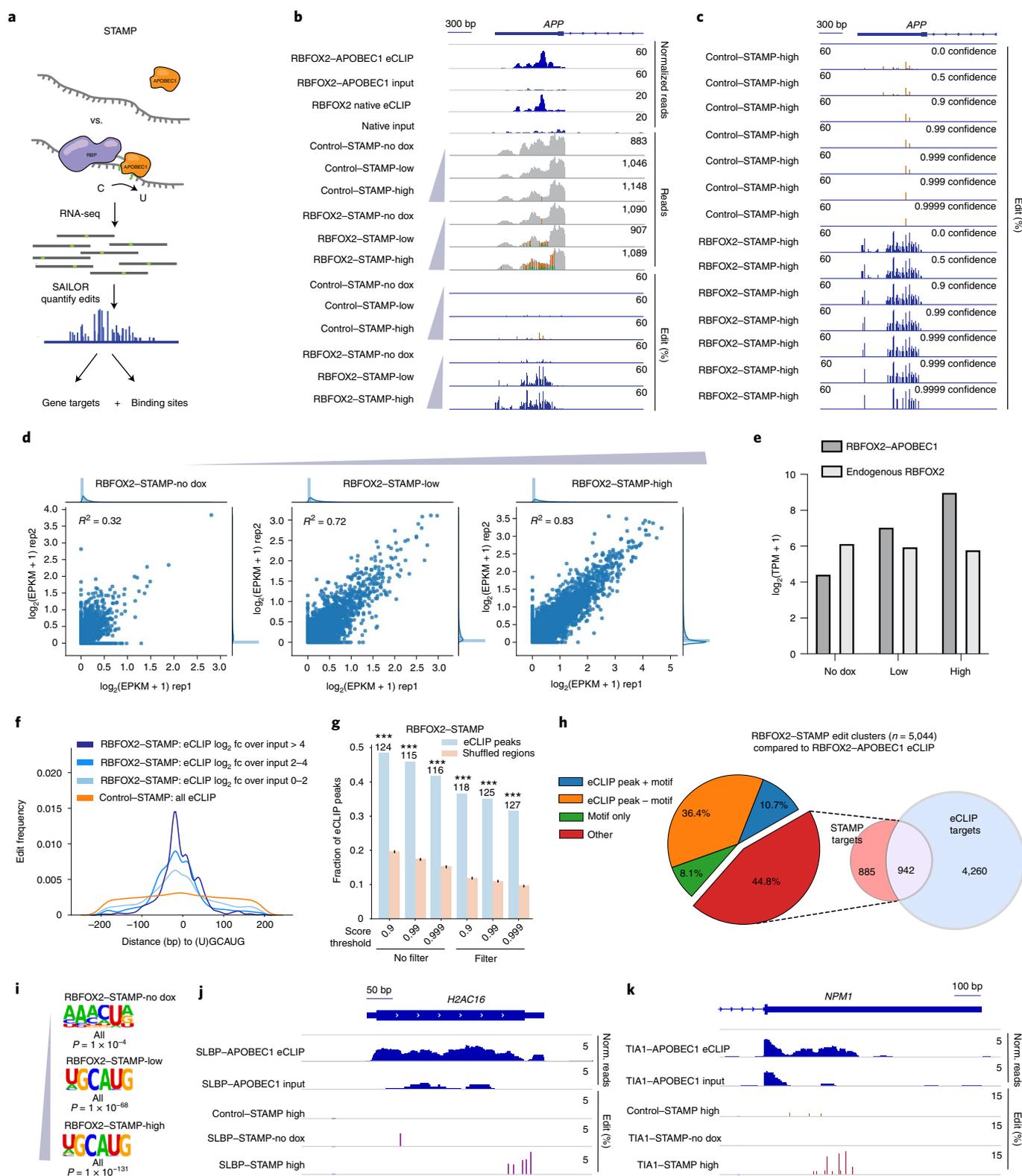
We next measured the nucleotide distance of RBFOX2–STAMP edits from the conserved RBFOX2 binding-site motif. For 2,852 RBFOX2 eCLIP peaks that harbor the canonical RBFOX2 motif (U) GCAUG, distances from the motif to RBFOX2–STAMP and control–STAMP (background) edits were determined within a 400-bp window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

Next, we developed a set of criteria that retrieves high-confidence edit clusters for RBP–STAMP while reducing false positives, analogous to peak calling in analyzing CLIP-seq datasets. We observed that the overlap of RBFOX2–STAMP edits with RBFOX2–APOBEC1 eCLIP peaks increased with increasing gene expression thresholds (Extended Data Fig. 1c), and we also anticipated more background edits within more highly expressed substrates. To minimize this background while enriching for true binding sites, we developed an edit-cluster-finding algorithm with gene-specific thresholds that assumed Poisson-distributed edit scores ϵ calculated for each site (Methods). Sites that satisfied gene-specific ϵ thresholds ($P < 0.05$ with adjusted Bonferroni correction for multiple-hypothesis testing) and SAILOR confidence-score thresholds were then merged with neighboring sites. Instances of edit sites with no neighboring edits within 100 bases in either direction were removed (workflow schematized in Extended Data Fig. 1d). These criteria established a set of 5,044 edit clusters for RBFOX2–STAMP (5.4% of the original unfiltered windows) and removed essentially all background control–STAMP sites (21 remaining, 0.04% of unfiltered windows; Extended Data Fig. 1e). Next, we determined the fraction of RBFOX2–APOBEC1 eCLIP peaks detected by these RBFOX2–STAMP edit clusters. We found that nearly half of all significant eCLIP peaks (\geq fourfold enriched over size-matched input; $P < 0.001$) overlapped with RBFOX2 edit clusters at a SAILOR confidence threshold of 0.9

Fig. 1 | RBP–STAMP edits mark specific RBP binding sites. **a**, 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-seq and quantified using our SAILOR analysis pipeline. **b**, Integrative genome viewer (IGV) browser tracks showing RBFOX2 and RBFOX2–APOBEC1 eCLIP peaks on the target gene *APP*, compared with control–STAMP and RBFOX2–STAMP signal and SAILOR-quantified edit fraction for increasing induction levels of fusions (doxycycline: 0 g ml⁻¹ (none), 50 ng ml⁻¹ (low) or 1 µg ml⁻¹ (high); 72 h). **c**, IGV tracks showing 72-h high-induction control–STAMP and RBFOX2–STAMP signal on the *APP* target gene at increasing confidence levels. **d**, RBFOX2–STAMP replicate correlations for the edited read counts per target normalized for length and coverage (EPKM). **e**, Quantification of expression from no doxycycline (0 ng ml⁻¹), low (50 ng ml⁻¹) or high (1 µg ml⁻¹) doxycycline induction of RBFOX2–APOBEC1 fusion compared to endogenous RBFOX2 expression. **f**, RBFOX2–STAMP and control–STAMP (background) edit frequency distribution within a 400-bp window flanking RBFOX2 eCLIP binding-site motifs, split into increasing levels of log₂ fold enrichment of eCLIP peak read density over size-matched input. **g**, Fraction of RBFOX2–APOBEC1 eCLIP peaks (log₂ fold change (FC) > 2 and $-\log_{10} P > 3$ over size-matched input) with RBFOX2–STAMP edit clusters, compared to size-matched shuffled regions, calculated at different edit-site confidence levels before and after site filtering (see Methods for filtering procedure). Numbers atop the bars are z-scores computed by comparing observed data with the distribution from random shuffles. Asterisks denote statistical significance; *** $P=0$, one-sided exact permutation test. **h**, Pie chart showing the proportion of filtered RBFOX2–STAMP edit clusters overlapping: (1) RBFOX2–APOBEC1 fusion high-confidence eCLIP peaks (log₂FC > 2 and $-\log_{10} P > 3$) containing the conserved RBFOX2 binding motif, (2) equally stringent eCLIP peaks not containing the conserved motif, (3) the conserved motif falling outside of eCLIP peaks or (4) neither eCLIP peaks nor conserved motifs. **i**, Motif enrichment using HOMER and shuffled background on RBFOX2–STAMP edit clusters for increasing RBFOX2–STAMP induction levels. **j**, IGV tracks showing control–STAMP and SLBP–STAMP edit fractions at no induction and high induction (doxycycline: 0 ng ml⁻¹ (none) or 1 µg ml⁻¹ (high); 72 h) on the target histone gene *H2AC16* compared to SLBP–APOBEC1 eCLIP. **k**, IGV tracks showing control–STAMP and TIA1–STAMP edit fractions at no induction and high induction (doxycycline: 0 ng ml⁻¹ (none) or 1 µg ml⁻¹ (high); 72 h) on the target gene *NPM1* compared to TIA1–APOBEC1 eCLIP.

for the edit sites, which was more than twofold higher compared to overlaps with randomly shuffled size-matched regions on exons of the same target genes (Fig. 1g). At higher SAILOR confidence thresholds, the fraction that overlaps decreased but the enrichment over background was preserved. We observed that 47% of RBFOX2–STAMP edit clusters overlapped with RBFOX2 eCLIP

peaks, irrespective of whether the eCLIP peaks contained known RBFOX2 binding motifs, and an additional 8% of the edit clusters contained the RBFOX2 motif (Fig. 1h). Interestingly, most clusters that did not overlap with eCLIP peaks were nevertheless located within eCLIP target genes at a distance from neighboring eCLIP peaks (Extended Data Fig. 1f). Subjecting control–STAMP



sites to our same criteria for RBFOX2–STAMP sites left essentially no background edit clusters to compare to eCLIP (Extended Data Fig. 1e). We also evaluated the orientation of the APOBEC1 fusion protein and observed that edit clusters and eCLIP peaks overlapped substantially from APOBEC1 fused to the N terminus of RBFOX2. However, the overlap was 20% smaller than what was observed for the C-terminal fusion, demonstrating that fusion orientation should be considered for each RBP of interest to maximize binding-site capture (Extended Data Fig. 1g). Lastly, we performed de novo motif discovery using high-confidence RBFOX2–STAMP edit clusters, assessing enrichment above a shuffled background for each gene region. These edit clusters were significantly enriched for the (U)GCAUG RBFOX2 binding motif, and the enrichments were correlated with the doxycycline dose and subsequent expression levels of RBFOX2–STAMP (Fig. 1i), demonstrating the sensitivity and specificity of STAMP for discovering RBP binding sites.

We next generated two additional HEK293T RBP–STAMP cell lines: one that inducibly expresses APOBEC1 fused to the histone stem-loop binding protein SLBP, and another that expresses a fusion to the stress granule protein TIA1 that binds target mRNA 3' UTRs^{3,34,35}. We noted similar STAMP-fusion expression levels compared to endogenous TIA1 and SLBP, as were observed for RBFOX2–STAMP (Extended Data Fig. 1h). As with RBFOX2–STAMP, we saw that the number of TIA1–STAMP edits on target genes increased with doxycycline concentration and were strongly correlated across replicates, with summary IDR analysis revealing thousands of reproducible edits that increased in number with increasing induction levels (Extended Data Fig. 1i). Comparison of SLBP–STAMP to SLBP–APOBEC1 eCLIP data showed that SLBP–STAMP edits were enriched compared to control–STAMP near eCLIP peaks within the 3' UTR of histone genes, such as *H2AC16* (Fig. 1j) adjacent to stem-loop regions, as expected. Comparison of control–STAMP and TIA1–STAMP to TIA1–APOBEC1 eCLIP revealed that there was inducible TIA1–STAMP edit enrichment overlapping the example eCLIP 3' UTR peak within the *NPM1* gene (Fig. 1k). Globally we found greater than 70% of all significantly reproduced SLBP eCLIP peaks⁶ (>fourfold enriched over size-matched input; $P < 0.001$, reproducible by IDR) overlapped with SLBP–STAMP edit clusters (Extended Data Fig. 1j), and more than 30% of all significant TIA1–APOBEC1 eCLIP peaks by the same criteria overlapped with TIA1–STAMP edit clusters (Extended Data Fig. 1k), with size-matched randomly shuffled regions on exons of the respective genes showing significantly lower concordance with edit clusters at any threshold for both RBPs. By de novo motif analysis, we also obtained the known eCLIP established U(A)-rich binding sequence from TIA1–STAMP edit clusters (Extended Data Fig. 1l). 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 edits are enriched in highly translated coding sequences and are responsive to mTOR inhibition.

Since ribosomes have extensive association with mRNAs during translation, we reasoned that ribosomal subunits fused to APOBEC1 (Ribo–STAMP) have the potential to edit mRNAs in a manner that reflects ribosome association. We generated independent HEK293T cell lines expressing APOBEC1 fusions to ribosomal subunits RPS2 and RPS3. For RPS2–STAMP and RPS3–STAMP, we observed that edits were enriched relative to control–STAMP on exons of protein-coding genes that are highly translated in HEK293T cells, such as *ATP5PB*³⁶, coincident with RPS3 eCLIP signal enrichment over size-matched input control (Fig. 2a). In comparison, RPS2–STAMP and RPS3–STAMP signals were minimally detected on highly expressed noncoding genes such as the long noncoding RNA *MALAT1*, which is localized to the cytoplasm in mitotic cell lines³⁷ (Fig. 2b). We performed replicate RPS2–STAMP

and control–STAMP inductions at low and high doxycycline concentrations for 24, 48 and 72 h and again observed dose-dependent STAMP-fusion expression compared to endogenous RPS2 levels (Extended Data Fig. 2a), with strong EPKM reproducibility between replicates ($R^2 = 0.6$ to 0.8), as well as low overlap (2.8% of all detectable edits) between control–STAMP and RPS2–STAMP edit sites at high induction (Extended Data Fig. 2b–d and Supplementary Table 4). As edits from RPS2–STAMP and RPS3–STAMP were present in coding sequence (CDS) regions and also in 3' UTR sequences (Fig. 2a), we needed to determine if these 3' UTR edits should be filtered or if they are coincident bystander edits. Comparison of EPKM values computed from CDS regions only, to EPKM values computed from both CDS and 3' UTR regions revealed a strong correlation, indicating that 3' UTR edits need not be excluded from downstream analyses and, in some instances, may provide edits otherwise missed if we considered only CDS regions in genes with short open-reading frames ($R^2 = 0.78$; Extended Data Fig. 2e).

To evaluate whether Ribo–STAMP can distinguish genes with varying levels of ribosome occupancy, we next compared combined genome-wide EPKM values from control–STAMP, RPS2–STAMP and RPS3–STAMP to RPKM values from ribosome-protected fragments (RPFs) obtained from standard ribosome profiling (Ribo-seq)³⁸ and to RPKM values from polyribosome-fraction-enriched RNA (polysome-seq)³⁹ experiments performed in HEK293 cells. For control–STAMP and for uninduced RPS2–STAMP, EPKM values were poorly correlated with Ribo-seq RPKM values ($R^2 = 0.32$ and $R^2 = 0.29$ respectively; Fig. 2c). At low and high levels of doxycycline induction, we found that the correlations between EPKM values for RPS2–STAMP and Ribo-seq RPF RPKM values improved substantially ($R^2 = 0.41$ and $R^2 = 0.46$, respectively; Fig. 2c). We observed a similar relationship when comparing RPS3–STAMP to Ribo-seq ($R^2 = 0.42$; Fig. 2d). RPS2–STAMP and polysome-seq measurements were also well correlated ($R^2 = 0.54$), consistent across replicates and improved at higher doxycycline induction concentrations and expression times (Fig. 2e and Extended Data Fig. 2f). As RPS2–STAMP had higher correlation with independent ribosome foot-printing approaches than RPS3–STAMP (Fig. 2d), we proceeded with RPS2–STAMP as the representative Ribo–STAMP fusion for downstream analysis. Meta-coding gene analysis of RPS2–STAMP edits for the top quartile of ribosome-occupied (Ribo-seq) genes revealed enrichment of edits within the CDS when compared to control–STAMP background edits and RBFOX2–STAMP edits, which showed the expected 3' UTR profile consistent with eCLIP (Fig. 2f). Enrichment of RPS2–STAMP edits within 3' UTRs likely indicates small-ribosomal subunit association with these accessible regions following ribosome translation termination by release factors, as we also observed 3' UTR signal from endogenous RPS3 eCLIP (Fig. 2a). These results are in agreement with previous studies revealing widespread 3' UTR ribosome footprints in both yeast and human cells^{40–42}. Together, these results demonstrate that Ribo–STAMP edit read counts track ribosome-occupancy measurements.

To determine if Ribo–STAMP edits detect translational perturbations, we performed stable high-induction RPS2–STAMP 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 kinase⁴³. Pharmacological inhibition of the mTOR pathway globally suppresses translation of mRNAs after initially suppressing translation of genes encoding the translational machinery itself⁴⁴. A 72-h Torin-1 treatment resulted in reproducible suppression in RPS2–STAMP edit distributions compared to vehicle-treated cells, exemplified by a marked decrease in edits on the top quartile of ribosome-occupied genes (Ribo-seq; Fig. 2g). RPS2–STAMP EPKM values were also significantly reduced upon Torin-1 treatment in the highest quartile of ribosome-occupied genes as defined

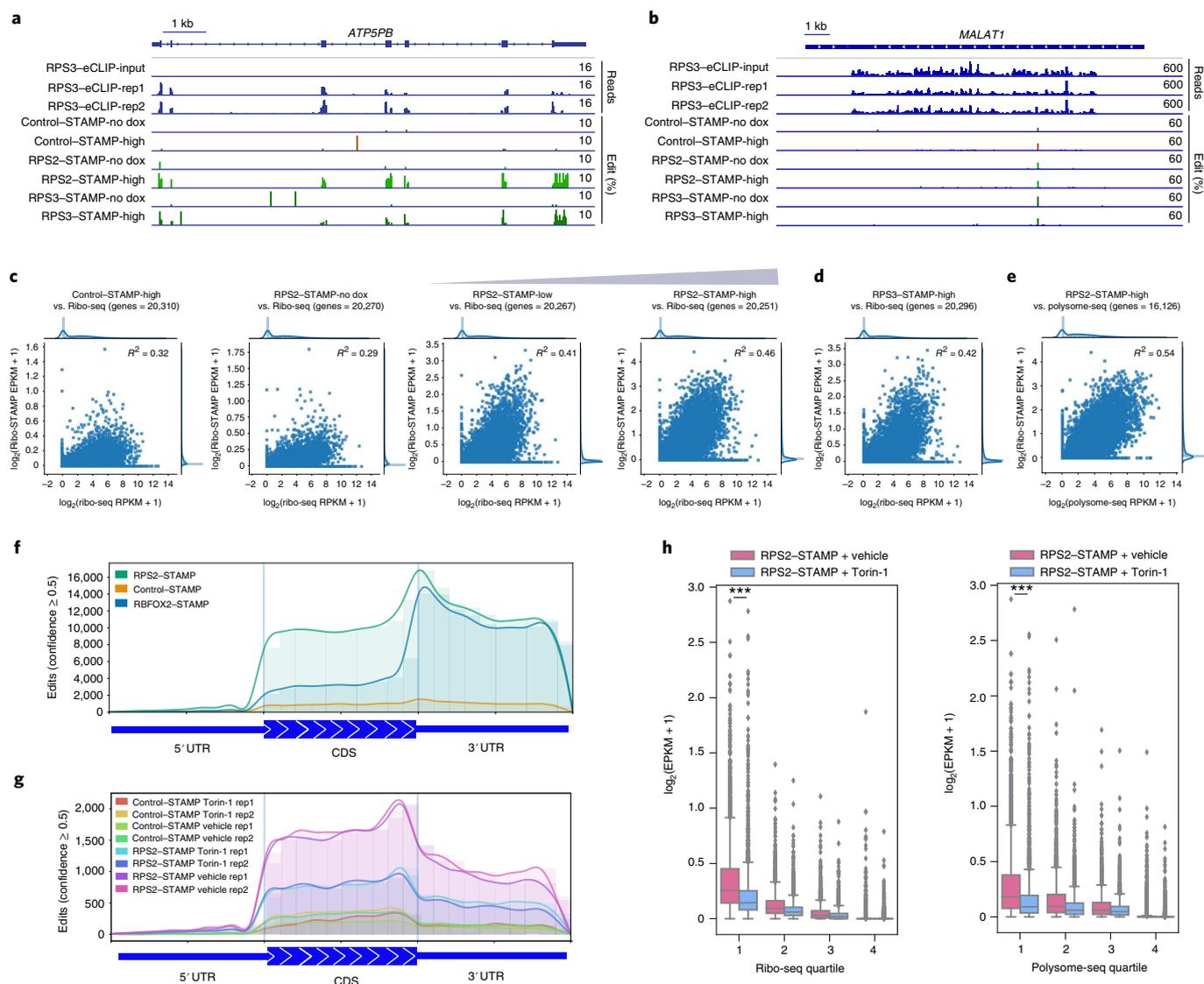


Fig. 2 | Ribo-STAMP edits mark highly translated coding sequences. a, IGV browser tracks displaying coding sequence edit frequency from control, RPS2-STAMP and RPS3-STAMP at no induction or 72-h high induction on the *ATP5PB* gene locus. RPS3 eCLIP and input reads are shown for comparison. **b**, IGV browser tracks as in **a** on the noncoding RNA *MALAT1*, showing no enrichment for RPS3 eCLIP reads, RPS2-STAMP edits or RPS3-STAMP edits. **c**, Genome-wide scatterplot comparison of control-STAMP and RPS2-STAMP EPKM values and Ribo-seq RPF RPKM values for increasing levels of RPS2-STAMP. **d**, Comparison as in **c** with Ribo-seq RPF RPKM values and EPKM values from RPS3-STAMP. **e**, Comparison as in **c** with polysome-seq RPKM values and EPKM values from RPS2-STAMP. **f**, Metagene plot showing edit (≥ 0.5 confidence score) distribution for high-induction RPS2-STAMP compared to control-STAMP and RBFOX2-STAMP across 5' UTR, CDS and 3' UTR gene regions for the top quartile ($n = 4,931$) of ribosome-occupied genes (Ribo-seq). **g**, Metagene plot as in **f** showing edit (≥ 0.5 confidence level) distribution for vehicle-treated 72-h high-induction RPS2-STAMP compared to replicate Torin-1-treated 72-h high-induction RPS2-STAMP across 5' UTR, CDS and 3' UTR gene regions for the top quartile of ribosome-occupied genes. **h**, Comparison of EPKM values from combined replicates ($n = 2$) vehicle-treated 72-h high-induction RPS2-STAMP compared to Torin-1-treated 72-h high-induction RPS2-STAMP showing significant signal reduction for top ribosome-occupied quartile genes containing Torin-1-sensitive TOP genes as detected by Ribo-seq (Q1, $P = 1.9 \times 10^{-147}$; $n = 3,589$ genes; Wilcoxon rank-sum one-sided test) and polysome profiling (Q1, $P = 7.7 \times 10^{-108}$; $n = 3,589$ genes; Wilcoxon rank-sum one-sided test).

by Ribo-seq (Q1, $P = 1.9 \times 10^{-147}$; Wilcoxon rank-sum test), and polysome-seq (Q1, $P = 7.7 \times 10^{-108}$; Wilcoxon rank-sum test), and all previously reported Torin-1-sensitive terminal oligopyrimidine (TOP) genes⁴⁴ were contained within these top quartiles (Fig. 2h). We observed no significant reduction in EPKM values for control-STAMP cells upon Torin-1 treatment for any matched comparisons (Extended Data Fig. 2g). Gene-level comparison of EPKM values for Torin-1 and vehicle-treated RPS2-STAMP on the highest quartile of ribosome-occupied genes as defined by Ribo-seq revealed

translation suppression from Torin-1 treatment (Extended Data Fig. 2h), with no corresponding difference in RPKM values between treated and untreated samples (Extended Data Fig. 2i). Together, these results demonstrate that dynamic translational responses are detected by Ribo-STAMP.

Long-read STAMP reveals isoform-specific binding profiles. To determine if STAMP enables RNA target detection on full-length mRNA isoforms using long-read sequencing technology, we

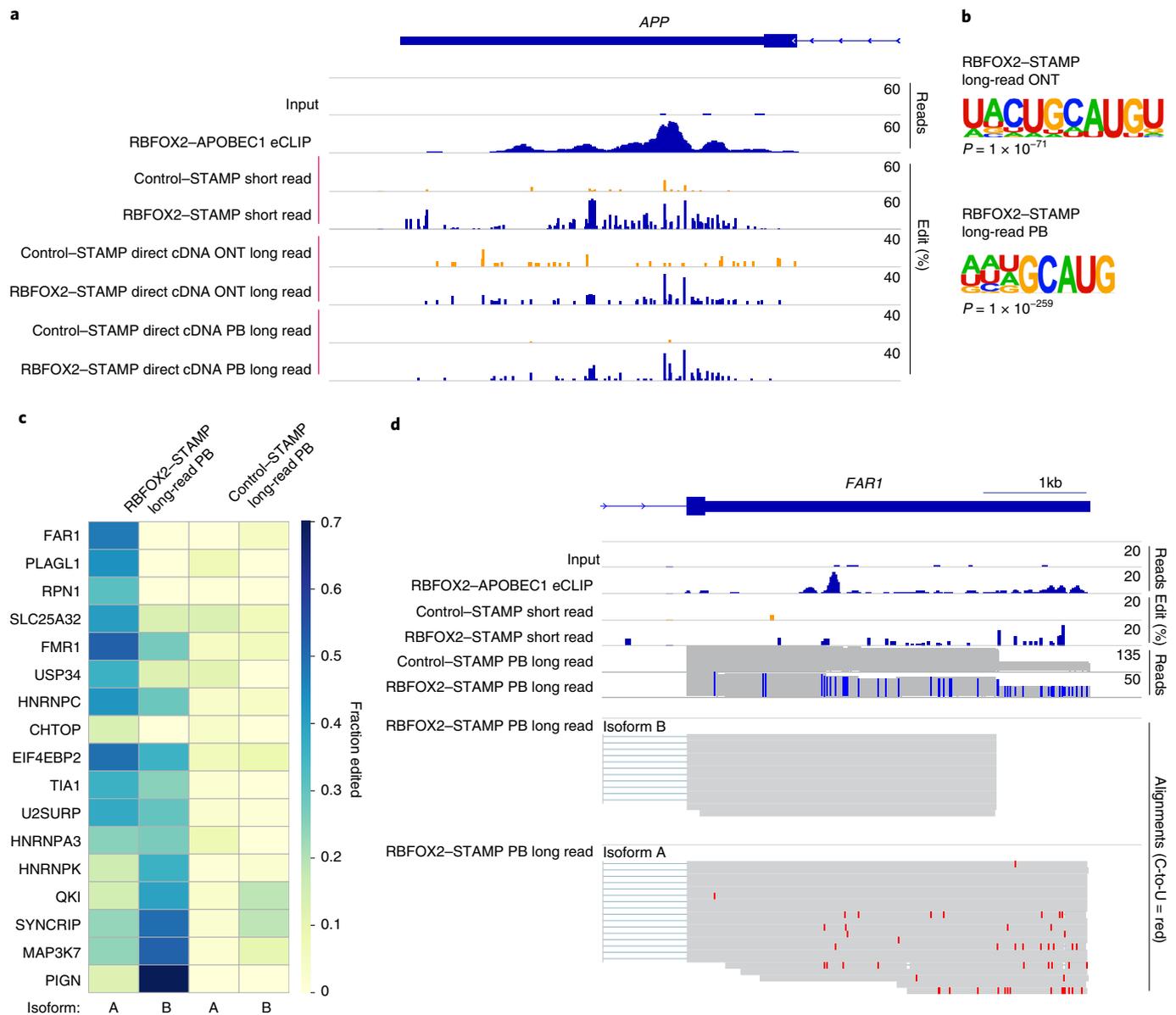


Fig. 3 | Long-read STAMP reveals isoform-specific binding profiles. a, IGV tracks showing RBFOX2 eCLIP peak on the target gene *APP*, compared with 72-h high-induction control–STAMP and RBFOX2–STAMP SAILOR-quantified edit fractions for both long-read (ONT or PB) direct cDNA, and short-read (next-generation sequencing) outputs. **b**, HOMER motif analysis of RBFOX2–STAMP long reads (ONT and PB) for edits above 0.99 confidence. **c**, Heat map of control–STAMP and RBFOX2–STAMP edit fractions on the two primary APA isoforms for the top differentially edited RBFOX2–STAMP APA targets. **d**, IGV tracks showing RBFOX2–APOBEC1 eCLIP peaks, control- and RBFOX2–STAMP short-read edit frequencies, and control- and RBFOX2–STAMP long-read (PB) alignments on the 2 primary isoforms of the target gene *FAR1*, with red colored C-to-U conversions on different isoforms.

performed 72-h stable high-induction RBFOX2–STAMP and control–STAMP and directly sequenced cDNA long reads with the Oxford Nanopore Technologies (ONT) and PacBio (PB) sequencing platforms^{45–47}. Both long-read sequencing approaches resulted in edit enrichment above control from RBFOX2–STAMP that overlapped with both eCLIP signal and short-read (Illumina) RBFOX2–STAMP signal, as illustrated by the target gene *APP* 3' UTR (Fig. 3a). Long-read high-confidence (≥ 0.99) RBFOX2–STAMP edits enriched the known RBFOX2 (U)GCAUG binding motif for both approaches (Fig. 3b). As PacBio has a lower base-calling error rate than Nanopore sequencing⁴⁸, we observed clear separation between control–STAMP and RBFOX2–STAMP edits and more significant motif extraction. We therefore focused on long-read

data from PacBio sequencing for downstream isoform-specific editing analysis.

To evaluate isoform-specific binding events, we calculated RBFOX2–STAMP or control–STAMP edit read fractions on the primary and secondary alternative polyadenylation (APA) isoforms of all genes (RBFOX2–STAMP, $n = 1,604$; control–STAMP, $n = 1,878$) that satisfied a minimal coverage threshold of ten reads per isoform for long reads obtained from PB sequencing. We observed differential isoform editing signatures for RBFOX2–STAMP compared to control–STAMP (Extended Data Fig. 3a and Supplementary Table 5). To illustrate, we displayed edits on the *FAR1* (Fig. 3c,d) and *PIGN* (Fig. 3c and Extended Data Fig. 3b) genes and observed RBFOX2–STAMP (but not control–STAMP) APA isoform-specific 3' UTR

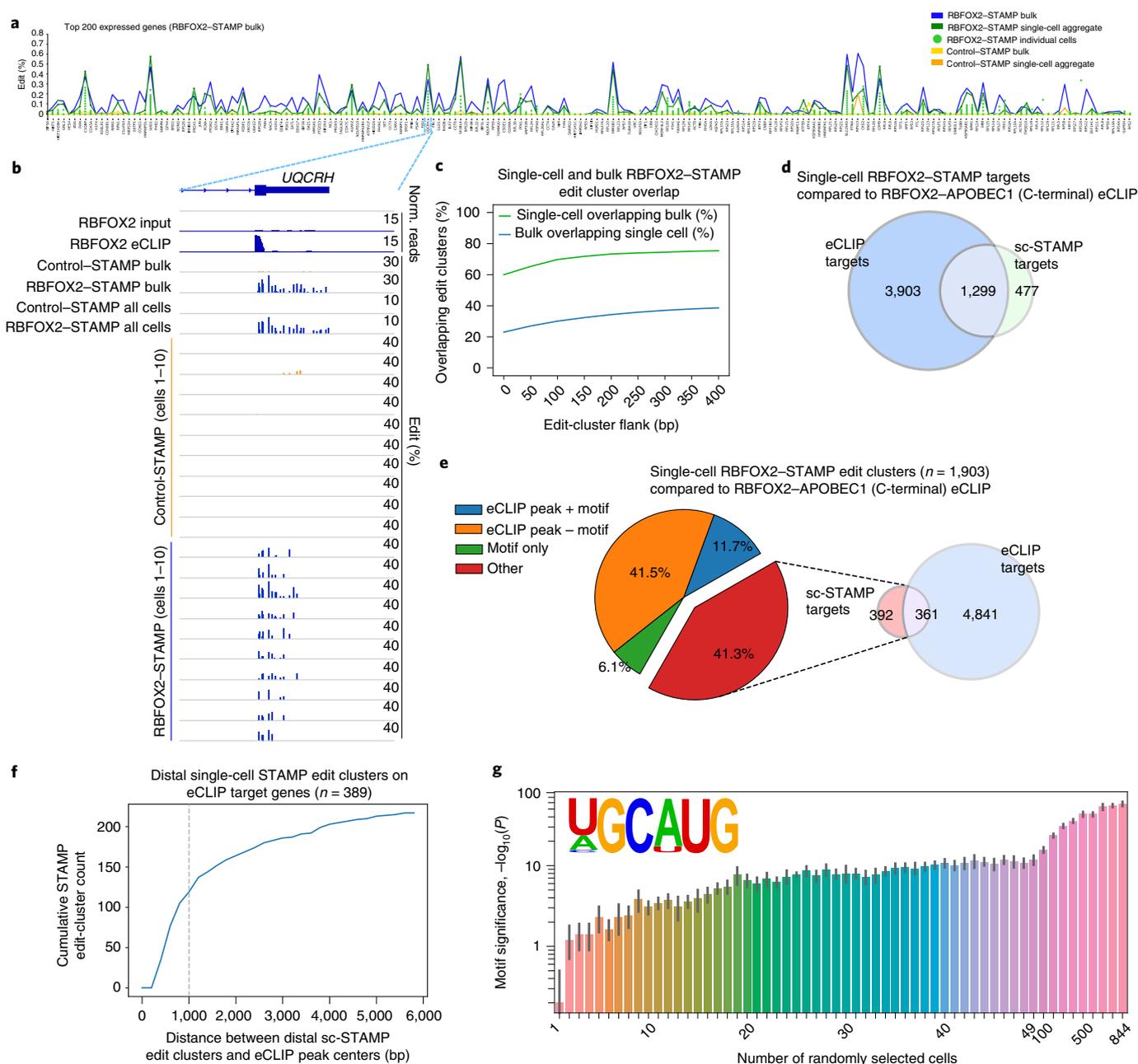


Fig. 4 | STAMP allows RBP binding-site detection at single-cell resolution. **a**, Edit fraction comparison of bulk 72-h high-induction control-STAMP and RBFOX2-STAMP with single-cell control-STAMP and RBFOX2-STAMP across the top 200 genes ranked by TPM values from bulk RBFOX2-STAMP RNA-seq. **b**, IGV tracks showing the RBFOX2 eCLIP peak on the target gene *UQCRH*, compared with RBFOX2-STAMP edit fractions for the top ten control-STAMP and RBFOX2-STAMP cells ranked by summed e scores. **c**, Evaluation of percentage overlap between bulk and single-cell edit clusters showing that 60–75% of single-cell edit clusters overlapped bulk edit clusters over increasing cluster-flanking regions. **d**, Overlap between RBFOX2-APOBEC1 eCLIP target transcripts (peaks $\log_2FC > 2$ and $-\log_{10}P > 3$ over input) and single-cell RBFOX2-STAMP edit clusters containing target transcripts. **e**, Pie chart showing the proportion of single-cell RBFOX2-STAMP edit clusters overlapping: (1) RBFOX2-APOBEC1 fusion high-confidence eCLIP peaks ($\log_2FC > 2$ and $-\log_{10}P > 3$ over input) containing the conserved RBFOX2 binding motif ((U)GCAUG), (2) equally stringent eCLIP peaks not containing the conserved motif, (3) the conserved motif falling outside of eCLIP peaks, or (4) neither eCLIP peaks nor conserved motifs. **f**, Cumulative distance measurement from single-cell RBFOX2-STAMP distal edit clusters to eCLIP peaks on targets genes. **g**, $-\log_{10}P$ 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 one-cell resolution.

edits, suggesting that RBFOX2 interacts with one of the isoforms but not the other. These isoform-specific binding sites coincided with both short-read RBFOX2-STAMP edit clusters and RBFOX2-APOBEC1 eCLIP peaks; however, the association of RBFOX2 to either isoform was indiscernible using these short-read approaches.

These results demonstrate that STAMP enables isoform-aware long-read detection of RBP-RNA interactions.

Detection of RBFOX2-RNA targets at single-cell resolution. To evaluate whether STAMP can discover RBP-RNA interactions in

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

Comparison of bulk and single-cell edit fractions for control-STAMP and RBFOX2-STAMP experiments across the top 200 expressed genes (ranked by transcripts per million (TPM) 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 (Fig. 4a). To illustrate, we next ranked individual control-STAMP and RBFOX2-STAMP cells by summed ϵ score and visualized edit fractions for the top ten cells on the RBFOX2 eCLIP target gene *UQCRH*. For all ten selected RBFOX2-STAMP cells, but not control-STAMP cells, we saw consistent edit signal in close proximity to the RBFOX2 eCLIP peak that overlapped with edit enrichment from both bulk RBFOX2-STAMP and the aggregate of all RBFOX2-STAMP cells (Fig. 4b), revealing that STAMP can define RBP binding sites at single-cell resolution. We saw strong concordance (80%) in the target genes that contained filtered high-confidence RBFOX2-STAMP edit clusters obtained from single-cell and bulk datasets (Extended Data Fig. 4a). At the binding-site level, 60% of these high-confidence single-cell edit clusters directly overlapped edit clusters obtained from bulk RBFOX2-STAMP, and ~70% of all single-cell edit clusters fell within 400bp of bulk edit clusters (Fig. 4c). In addition, we found that 73% of single-cell STAMP targets that contained edit clusters also contained significant RBFOX2-APOBEC1 eCLIP peaks ($P < 0.001$; Fig. 4d). As with bulk RBFOX2-STAMP, a majority of the single-cell RBFOX2-STAMP edit clusters overlapped eCLIP peaks and harbored RBFOX2 binding motifs (Fig. 4e), with a large number of clusters that did not directly overlap eCLIP peaks still present in target genes generally within 1,000bp of the neighboring eCLIP peak (Fig. 4f). As expected, the single-cell RBFOX2-STAMP eCLIP peak capture rate was associated with target expression level (Extended Data Fig. 4b). De novo motif analysis from edit clusters by randomly downsampling the numbers of single cells analyzed identified the canonical (U)GCAUG motif with significance, even to the resolution of one cell (Fig. 4g), showcasing the strength of single-cell STAMP.

Deconvolution of RBP-specific and cell type-specific RNA binding. The ability of STAMP to recover RBP-RNA targets in single cells suggests that targets of multiple RBPs can be simultaneously discovered from a single multiplexed experiment. In our RBFOX2-STAMP experiment, we separately performed 72-h high-induction TIA1-STAMP, before mixing equal numbers of RBFOX2-STAMP and TIA1-STAMP cells, followed by scRNA-seq. Cells harboring capture sequences for TIA1-STAMP and RBFOX2-STAMP were better distinguished from each other and from control-STAMP cells by uniform manifold approximation and projection (UMAP)

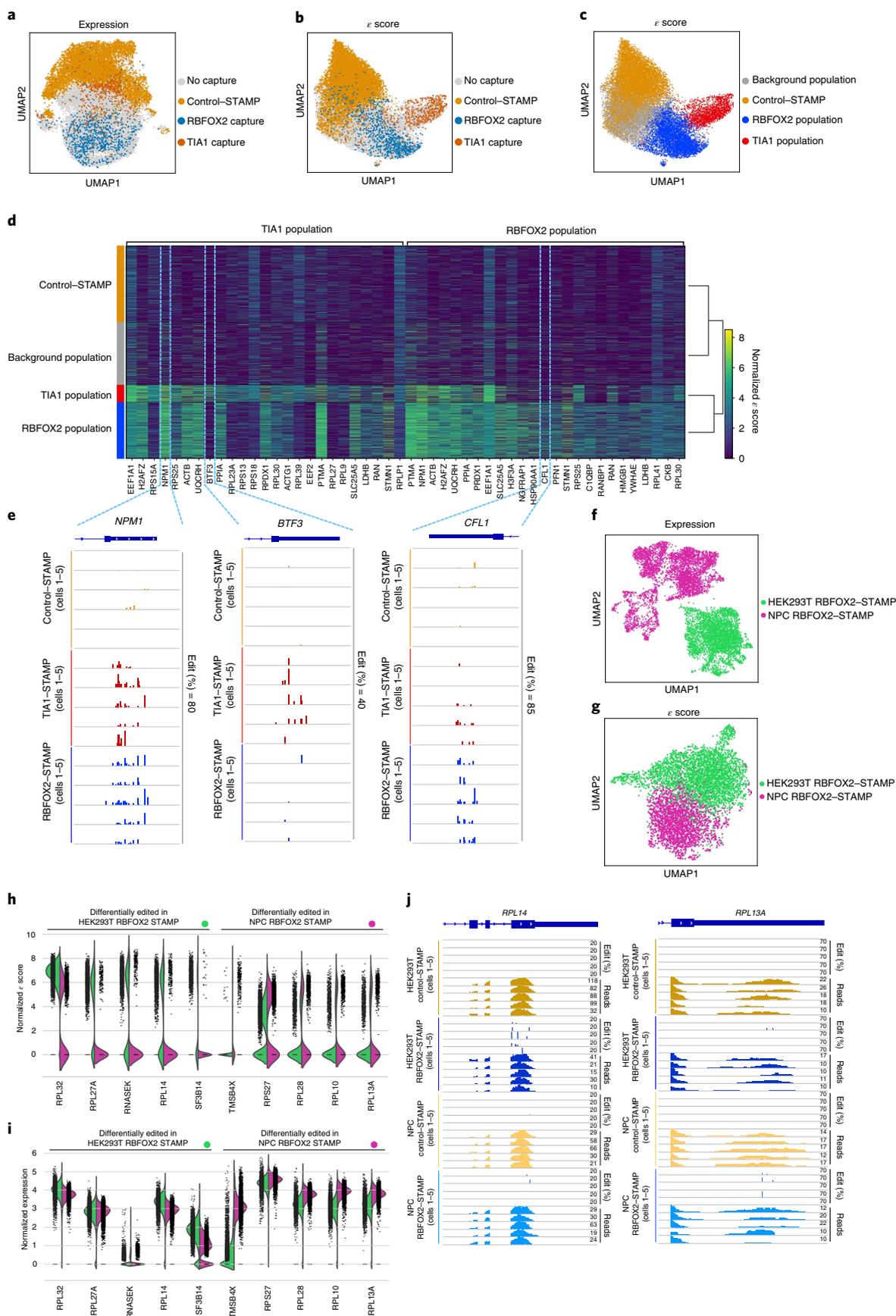
visualization using ϵ scores, than by gene expression (Fig. 5a,b and Extended Data Fig. 5a), congruent with our expectations that the single-cell ϵ score profiles of TIA1-STAMP and RBFOX2-STAMP targets were sufficiently distinct. UMAP visualization of ϵ scores further revealed that control-STAMP cells ($n=8,117$ cells) were distinct from RBFOX2-STAMP and TIA1-STAMP 'capture cells' (Fig. 5b). Using Louvain clustering by ϵ score profiles, we thus defined an RBFOX2 population ($n=6,003$ cells), a TIA1 population ($n=1,841$ cells) and a background population ($n=6,623$ cells) for further analysis (Extended Data Fig. 5b). Overlap with control (Fig. 5c) and reclustering in the expression space (Extended Data Fig. 5c) 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 (Extended Data Fig. 5d), and TIA1 and RBFOX2 clusters displayed distinct editing profiles when compared to control-STAMP (Fig. 5d and Supplementary Table 5). We ranked cells based on summed ϵ scores to select cells with the most robust editing and found that the top five cells for each RBP displayed edit enrichment on the shared RBFOX2-STAMP and TIA1-STAMP target *NPML1*, which was also detected as a TIA1 target by eCLIP and bulk TIA1-STAMP (Fig. 1k). Edit enrichments for individual cells were specific to TIA1-STAMP on the *BTF3* target gene, and to RBFOX2-STAMP on the *CFL1* target gene (Fig. 5e), demonstrating that the targets and binding sites of multiplexed RBP-STAMP fusions can be delineated from edit signatures within single-cell experiments.

To identify cell type-specific RBP targets using single-cell STAMP, we performed STAMP in HEK293T cells and pluripotent stem cell-derived neural progenitor cells (NPCs) by transient transfection with plasmids constitutively expressing either RBFOX2-STAMP 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 (Fig. 5f, Extended Data Fig. 5e and Supplementary Table 6). UMAP clustering on ϵ score also resulted in separation of cell types (as determined by gene expression clustering) based on RBFOX2-STAMP edits (Fig. 5g and Supplementary Table 7), and we extracted the RBFOX2 binding motif using edit clusters from 2,178 NPCs editing 468 target genes, and 3,258 HEK293 cells editing 939 target genes (Extended Data Fig. 5f). Analysis of the top RBFOX2-STAMP differentially edited genes between cell types revealed cell type-specific targets (Fig. 5h) that were often not differentially expressed (Fig. 5i), indicating cell type-specific RNA-protein interactions independent of target expression levels. Individual cell edits for the top five control-STAMP 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 NPCs such as *RPL13A* (Fig. 5j). Together, these results indicate that cell type-specific targets and binding sites can be extracted from RBFOX2-STAMP edit signatures by scRNA-seq within a mixture of heterogeneous cell types.

Fig. 5 | Deconvolution of multiple RBPs and cell type-specific targets. **a**, UMAP analysis of gene expression from merged 72-h high-induction control-STAMP and RBFOX2:TIA1-STAMP cells with capture sequence RBFOX2-STAMP (blue; $n=844$) and TIA1-STAMP cells (red; $n=527$) highlighted. **b**, UMAP analysis using ϵ scores rather than gene expression after merging 72-h high-induction control-STAMP cells (orange). **c**, UMAP plot as in **b** color coded by ϵ score Louvain clustering into RBFOX2-population (blue), TIA1-population (red) and background-population (gray) populations with control-STAMP cells (orange) overlaid. **d**, Heat map of normalized ϵ score signatures for RBFOX2 and TIA1-population cells compared to control-STAMP and background cells on the top 25 differentially edited gene targets. **e**, IGV browser tracks showing SAILOR-quantified edit fractions for the top five control-STAMP, RBFOX2-STAMP and TIA1-STAMP cells (ranked by summed ϵ scores) on the *NPML1*, *BTF3* and *CFL1* gene targets. **f**, UMAP analysis of merged 72-h high-induction RBFOX2-STAMP mixed NPCs and HEK293T cells clustered by expression. **g**, UMAP analysis as in **f** using ϵ scores. **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 five control-STAMP and RBFOX2-STAMP cells (ranked by summed ϵ scores) on the *RPL14* and *RPL13A* gene targets.

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-h

high-induction control-STAMP and RPS2-STAMP and conducted scRNA-seq. To distinguish control-STAMP and RPS2-STAMP cell populations, we computed EPKM measurements for protein-coding



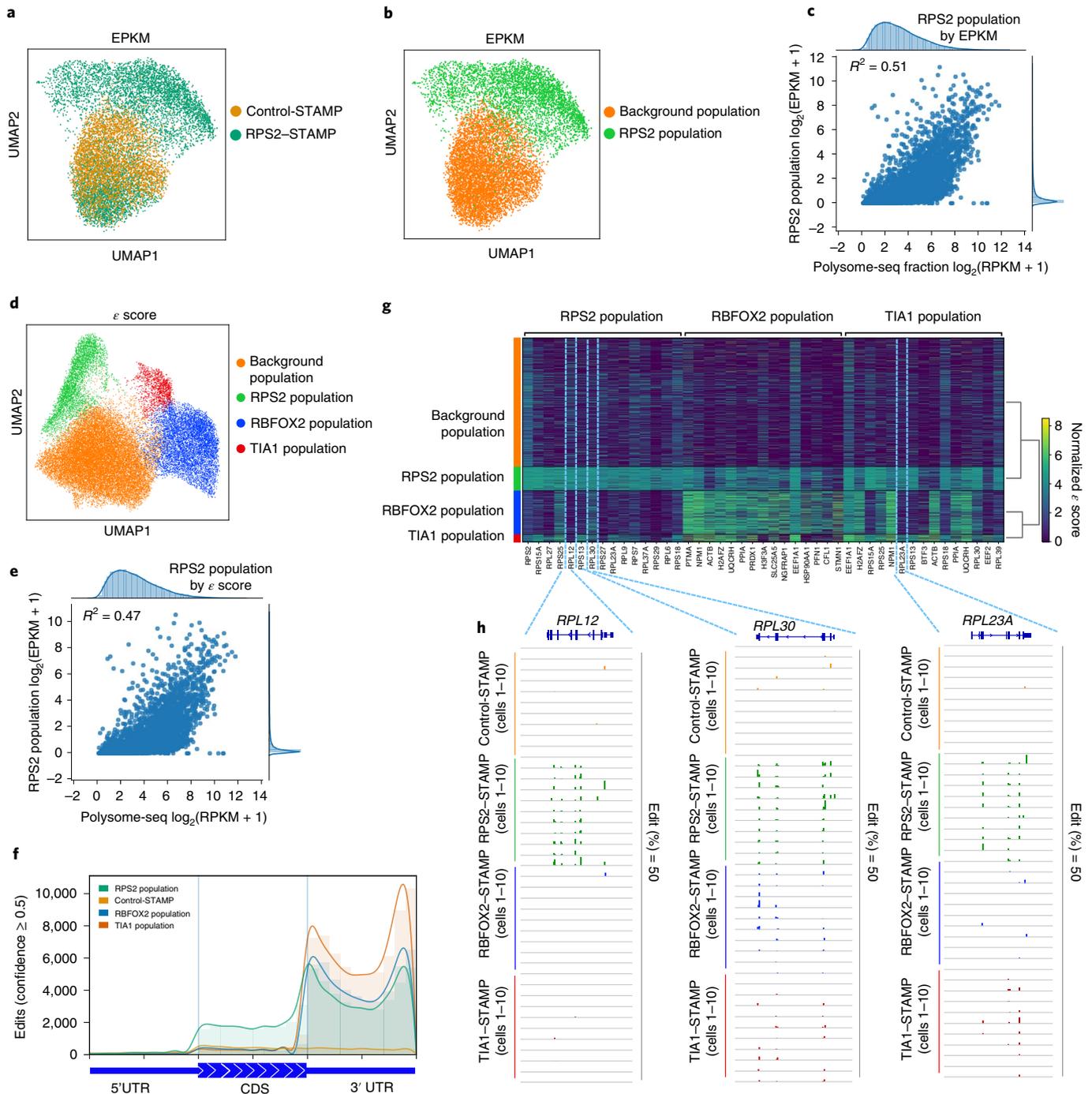


Fig. 6 | Ribo-STAMP reveals ribosome occupancy from individual cells. **a**, UMAP analysis of EPKM values for 72-h high-induction RPS2-STAMP (green) and control-STAMP (orange). **b**, UMAP analysis of cells shown in **a** with EPKM Louvain clustering into background-population and RPS2-population cells. **c**, Comparison of EPKM-derived RPS2-population CDS and 3' UTR EPKM values with polysome-seq RPKM values. **d**, UMAP plot color coded by ϵ score Louvain clustering into background-cluster (orange), RBFOX2-cluster (blue), TIA1-cluster (red) and 677 RPS2-cluster (green) cells from merged 72-h high-induction STAMP experiments. **e**, Comparison of ϵ score-derived RPS2-population CDS and 3' UTR EPKM values with polysome-seq RPKM values. **f**, Metagene plot showing distribution for aggregate cell edits (≥ 0.5 confidence level) from control-STAMP, RPS2-cluster, TIA1-cluster and RBFOX2-cluster cells across 5' UTR, CDS and 3' UTR gene regions for the top quartile of ribosome-occupied genes. **g**, Heat map of normalized ϵ score signatures for RPS2-population, RBFOX2-population and TIA1-population cells compared to background cells on the top 15 differentially edited gene targets. **h**, IGV browser tracks showing edit fractions for the top ten control-, RPS2-STAMP, RBFOX2-STAMP and TIA1-STAMP cells (ranked by summed ϵ scores) on the *RPL12*, *RPL30* and *RPL23A* gene targets.

genes for each cell. EPKM-based UMAP representation (Fig. 6a) followed by Louvain clustering (Fig. 6b) revealed a group of RPS2-STAMP (RPS2-population) cells that was clearly distinct from a

population of background cells that contained a mixture of both control-STAMP and RPS2-STAMP cells (background population). Focusing on this RPS2 population, we showed that EPKM

values (from CDS and 3' UTR) aggregated from the 3,917 single cells correlated meaningfully ($R^2=0.53$) with genome-wide EPKM values from bulk Ribo-STAMP (Extended Data Fig. 6a). We note that EPKM values computed from edits within the combination of CDS and 3' UTR regions, compared to only CDS regions, correlated strongly ($R^2=0.81$; Extended Data Fig. 6b); therefore, we included 3' UTR-derived edit measurements. We next addressed if aggregated single-cell Ribo-STAMP EPKM values can approximate ribosome-occupancy measurements derived from bulk polyribosome-fraction-enriched RNA (polysome-seq). We first assessed if RNA abundance measurements for total mRNA from Ribo-STAMP and polysome-seq experiments were in good agreement and observed a positive relationship ($R^2=0.54$; Extended Data Fig. 6c). We then compared Ribo-STAMP mRNA edits to polysome-seq mRNA abundance and observed less agreement between these measurements ($R^2=0.32$; Extended Data Fig. 6d), suggesting that Ribo-STAMP edit enrichment is not simply dictated by transcript levels. In contrast, polyribosome-fraction-enriched RNA measurements from polysome-seq were well correlated with Ribo-STAMP edits ($R^2=0.51$; Fig. 6c), implying that single-cell Ribo-STAMP edit enrichments are more closely associated with ribosome occupancy than with transcript abundance. These results strongly indicate that single-cell Ribo-STAMP, like single-cell RBP-STAMP, recapitulates results from bulk experiments and correlates well with standard measurements from orthogonal bulk approaches.

Inspired by our ability to define RBP-specific populations from RBP-STAMP cell mixtures using editing information alone (Fig. 5b,c), we next integrated Ribo-STAMP with RBP-STAMP to define ribosome association and RBP binding sites in parallel after merging all control-STAMP, RBFOX2-STAMP, TIA1-STAMP and RPS2-STAMP single-cell edit matrices. UMAP visualization of single-cell, transcriptome-wide ϵ scores revealed that control-STAMP cells overlapped with a subpopulation of RBFOX2-STAMP, TIA1-STAMP and RPS2-STAMP cells (Extended Data Fig. 6e), highlighting cells that have similar background-level edit patterns. Louvain clustering within the UMAP projection space defined four distinct groups of single cells for downstream analysis: (1) RPS2-population cells ($n=3,621$ cells), (2) RBFOX2-population cells ($n=7,000$ cells), containing the majority (92%) of RBFOX2 cells identified by capture sequencing, (3) TIA1-population cells ($n=1,312$ cells), containing the majority (57%) of TIA1 capture cells and (4) a background population ($n=20,655$ cells), composed of control-STAMP cells and any cells that overlapped spatially with control-STAMP cells (Fig. 6d and Extended Data Fig. 6f). The ϵ score-derived RPS2 population was 90% matched to the EPKM-derived RPS2 population (Extended Data Fig. 6g) and also had good EPKM value correlation with polysome-seq measurements (Fig. 6e). Metagene plotting of edits from these four subgroups for the top quartile of ribosome-occupied genes (Ribo-seq; $n=4,931$ genes) demonstrated CDS enrichment for single-cell RPS2-STAMP edits compared to more 3' UTR-centric enrichment for single-cell RBFOX2-STAMP and TIA1-STAMP (Fig. 6f), in agreement with our results from bulk experiments (Fig. 2f). Differential ϵ score analysis showed distinct editing signatures for RPS2-population, RBFOX2-population and TIA1-population cells compared to the background population (Fig. 6g and Supplementary Table 8). To illustrate, the top ten cells ranked by summed ϵ scores exhibited the expected specific editing signatures on the *RPL12*, *RPL30* and *RPL23A* target transcripts (Fig. 6h). These results highlight the capacity of STAMP to reveal RBP targets and ribosome association in parallel at single-cell resolution.

Discussion

We have developed an experimental and computational workflow called STAMP, which allows antibody-free detection of RBP and ribosome interactomes (Ribo-STAMP) by standard RNA-seq and

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

STAMP has distinct advantages over TRIBE, as TRIBE generally yields only gene-level target information and not binding sites, with one to two edits on average detectable in any given target^{17,20,22,23}. The sparse editing signal by ADAR deaminase domains is due to the preference for ADAR to edit double-stranded RNAs that contain a bulged mismatch²⁴, an infrequent occurrence on single-stranded mRNAs transcriptome wide^{24,49}. In contrast, APOBEC enzymes access cytosines in single-stranded RNA that constitute ~25–35% of nucleotides in any given mammalian transcript and produce clusters of edits (between 10 and 1,000 edits at target sites). In addition, structured RNA is reduced from coding regions by active translation⁵⁰, making ribosome interactions that are easily detectable by Ribo-STAMP not feasible with ADAR-fusion approaches. Indeed, the RBP FMR1 fusions to ADAR, which were expected to be very frequent in the coding regions of genes, resulted in only four confident edits across the ~15-kb coding region of the showcased *POE* gene²⁰. The higher likelihood of encountering APOBEC1 cytosine substrates within single-stranded mRNA enables STAMP-mediated discovery of RBP-RNA sites with such high sensitivity and specificity that de novo discovery of conserved binding-site motifs can be extracted from even one single cell.

Antibody-based methodologies such as CLIP and RIP are staples used to identify RNA-binding sites and targets of RBPs. Our STAMP approach offers several advantages. First, CLIP is generally constrained by input requirements, frequently needing thousands to millions of cells. Here we demonstrate that STAMP can be used reliably at single-cell resolution to identify RNA targets, binding sites and even extract motifs from a few cells to a single cell. STAMP enables the combined identification of RBP binding sites and global measurement of gene expression, a long-standing goal for the gene expression, genomics and RNA communities. Second, CLIP requires fragmentation to separate bound and unbound RNA, but that precludes the discovery of isoform-dependent binding sites on mRNAs that may differ by an exon or translated regions. We show here that STAMP allows long-read assessment to distinguish RBP binding on different transcript isoforms. Further, direct RNA-seq has recently been demonstrated to be RNA-modification sensitive⁵¹, which opens the possibility of using STAMP to detect modification-sensitive RNA-protein interactions.

In our study, we utilized poly(A)⁺ mRNA-seq (other than total RNA-seq for SLBP-STAMP) to characterize binding interactions for RBFOX2, TIA1, RPS2 and RPS3. However, aside from their

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

Currently, antibody-free methods like STAMP and TRIBE require fusion of the protein of interest to a modifying enzyme, which may not be feasible for all RBPs. In addition, the editing on this time frame may have unintended consequences depending on the protein of interest. Our current version of Ribo–STAMP yields detectable edits within 12–24 h, a timeframe that may dampen the capacity to detect rapid translational responses and may lead to unintended expression modulations due to recoding of transcripts and the possible introduction of nonsense or frameshift mutations. Therefore, it is important to consider the duration of expression of Ribo–STAMP as it relates to the dynamics being assessed and to downstream unintended perturbations. Extended Ribo–STAMP expression could also explain the somewhat unexpected 3' UTR edit enrichment that we observed, although 3' UTR enrichment of edits appears to be a generalized phenomenon for both TRIBE/HyperTRIBE and DART-seq approaches, likely due to editing modules accessing susceptible 3' UTR sequence elements distal to actual fusion binding sites^{19,22,26}. Although we demonstrate that this 3' UTR Ribo–STAMP editing is inconsequential to ribosome-occupancy correlations with other profiling methods, these edits may nonetheless be biologically relevant. Gold-standard ribosome profiling (Ribo-seq) with alternative digestion conditions in human cells has uncovered widespread 3' UTR ribosome footprinting⁴², and translation complex profile sequencing in both yeast⁴¹ and human cells⁴⁰ revealed small-ribosomal-subunit-specific enrichment within 3' UTRs, likely attributable to ribosomal recycling-mediated mRNA interactions. In the future, we anticipate that engineering of fusion orientation, in addition to fine-tuning STAMP expression levels and duration of the overexpression window, will be useful to obtain editing profiles that are maximally informative for different RBPs and ribosomal subunits. For each new use case of the current version of the STAMP approach, we recommend testing edit signature responses to both fusion orientation and fusion expression levels using short-read RNA-seq from bulk cells and comparing these to gold-standard orthogonal methods such as CLIP and Ribo-seq before proceeding to long-read and single-cell resolution applications.

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

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-021-01128-0>.

Received: 2 October 2020; Accepted: 26 March 2021;

Published online: 7 May 2021

References

- Singh, G. et al. The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.* **84**, 325–354 (2015).
- Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
- Van Nostrand, E. L. et al. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA-binding proteins. *Genome Biol.* **21**, 90 (2020).
- Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA–protein interactions. *Nat. Methods* **16**, 225–234 (2019).
- Wheeler, E. C., E. L. Van Nostrand, E. L. & Yeo, G. W. Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdiscip. Rev. RNA* **9**, e1436 (2018).
- Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
- Perez-Perri, J. I. et al. Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun.* **9**, 4408 (2018).
- Calviello, L. & Ohler, U. Beyond read-counts: Ribo-seq data analysis to understand the functions of the transcriptome. *Trends Genet.* **33**, 728–744 (2017).
- Ingolia, N. T. et al. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
- Lee, F. C. Y. & Ule, J. Advances in CLIP technologies for studies of protein–RNA interactions. *Mol. Cell* **69**, 354–369 (2018).
- Clamer, M. et al. Active ribosome profiling with ribolace. *Cell Rep.* **25**, 1097–1108 (2018).
- 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* **10**, 1213–1218 (2013).
- Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA-sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 1–14 (2018).
- Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
- Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
- Shahi, P. et al. Abseq: ultrahigh-throughput single-cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* **7**, 44447 (2017).
- Nguyen, D. T. T. et al. HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nat. Commun.* **11**, 2026 (2020).
- Medina-Munoz, H. C. et al. Records of RNA locations in living yeast revealed through covalent marks. *Proc. Natl Acad. Sci. USA* **117**, 23539–23547 (2020).
- Jin, H. et al. TRIBE editing reveals specific mRNA targets of eIF4E-BP in *Drosophila* and in mammals. *Sci. Adv.* **6**, eabb8771 (2020).
- McMahon, A. C. et al. TRIBE: hijacking an RNA-editing enzyme to identify cell-specific targets of RNA-binding proteins. *Cell* **165**, 742–753 (2016).
- Lapointe, C. P. et al. Protein–RNA networks revealed through covalent RNA marks. *Nat. Methods* **12**, 1163–1170 (2015).
- Xu, W., Rahman, R. & Rosbash, M. Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. *RNA* **24**, 173–182 (2018).
- Rahman, R. et al. Identification of RNA-binding protein targets with HyperTRIBE. *Nat. Protoc.* **13**, 1829–1849 (2018).
- 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.* **23**, 426–433 (2016).
- Navaratnam, N. et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J. Biol. Chem.* **268**, 20709–20712 (1993).
- Meyer, K. D. DART-seq: an antibody-free method for global m⁶A detection. *Nat. Methods* **16**, 1275–1280 (2019).
- Deffit, S. N. et al. The *C. elegans* neural editome reveals an ADAR target mRNA required for proper chemotaxis. *Elife* **6**, e28625 (2017).

28. 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.* **6**, 599–607 (2014).
29. Lovci, M. T. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* **20**, 1434–1442 (2013).
30. 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.* **16**, 130–137 (2009).
31. 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.* **281**, 12468–12474 (2006).
32. Van Nostrand, E. L. et al. CRISPR–Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA-binding proteins. *Methods* **118–119**, 50–59 (2017).
33. Li, Q. H. et al. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
34. Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843–854 (2008).
35. Gilks, N. et al. Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol. Biol. Cell* **15**, 5383–5398 (2004).
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. USA* **115**, E9325–E9332 (2018).
37. Yang, F. et al. MALAT-1 interacts with hnRNP C in cell cycle regulation. *FEBS Lett.* **587**, 3175–3181 (2013).
38. Zhang, P. et al. Genome-wide identification and differential analysis of translational initiation. *Nat. Commun.* **8**, 1749 (2017).
39. Tan, F. E. et al. A transcriptome-wide translational program defined by LIN28B expression level. *Mol. Cell* **73**, 304–313 (2019).
40. Wagner, S. et al. Selective translation complex profiling reveals staged initiation and co-translational assembly of initiation factor complexes. *Mol. Cell* **79**, 546–560 (2020).
41. Archer, S. K. et al. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* **535**, 570–574 (2016).
42. Miettinen, T. P. & Bjorklund, M. Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res.* **43**, 1019–1034 (2015).
43. Thoreen, C. C. et al. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J. Biol. Chem.* **284**, 8023–8032 (2009).
44. Thoreen, C. C. et al. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
45. Jain, M. et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
46. Ardui, S. et al. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res.* **46**, 2159–2168 (2018).
47. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
48. Fu, S., Wang, A. & Au, K. F. A comparative evaluation of hybrid error correction methods for error-prone long reads. *Genome Biol.* **20**, 26 (2019).
49. Song, Y. et al. irCLASH reveals RNA substrates recognized by human ADARs. *Nat. Struct. Mol. Biol.* **27**, 351–362 (2020).
50. Beaudoin, J. D. et al. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* **25**, 677–686 (2018).
51. Lorenz, D. A. et al. Direct RNA sequencing enables m⁶A detection in endogenous transcript isoforms at base-specific resolution. *RNA* **26**, 19–28 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Plasmid construction. For the generation of stable cell lines, all RBP-STAMP mammalian expression constructs were in one of two lentiviral Gateway (Invitrogen) destination vector backbones: (1) pLIX403_APOBEC_HA_P2A_mRuby or (2) pLIX403_Capture1_APOBEC_HA_P2A_mRuby. pLIX403_APOBEC_HA_P2A_mRuby was cloned by amplification (Cloneamp, Takara Bio) of APOBEC1_HA_P2A cassette after removal of the YTH cassette from APOBEC1-YTH (a gift from K. 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 D. Root; Addgene plasmid no. 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) containing 10x Feature Barcode Capture Sequence 1 with Gibson assembly reaction into MluI digested backbone pLIX403_APOBEC_HA_P2A_mRuby in frame and immediately upstream of the APOBEC1 ORF. RBP open-reading frames (ORFs) were obtained from human Orfeome 8.1 (2016 release) donor plasmids (pDONR223) when available, or amplified (Cloneamp, Takara Bio) from cDNA obtained by SuperScript III (Invitrogen) RT-PCR of HEK293XT cell purified RNA (Direct-zol, Zymogen) and inserted into pDONR223 by Gateway BP Clonase II reactions (Invitrogen). Donor ORFs were inserted in frame upstream of APOBEC1 or Capture Sequence 1 APOBEC1 by gateway LR Clonase II reactions (Invitrogen). For transient transfections of HEK293T cells and NPCs, constructs were modified from pCMV-BE1-YTH-HA plasmid (a gift from K. 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 ORF 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⁻¹ D-glucose) supplemented with 10% FBS (Gibco) at 37°C with 5% CO₂. Cells were periodically passaged once at 70–90% confluency by dissociating with TrypLE Express Enzyme (Gibco) at a ratio of 1:10. The stable HEK293XT cell lines RBFox2-STAMP, TIA1-STAMP, SLBP-STAMP, RPS2-STAMP, RPS3-STAMP and control-STAMP were generated as described below by transducing ~1 million cells with 8 µg ml⁻¹ polybrene and 1 ml viral supernatant in DMEM + 10% FBS at 37°C for 24 h, followed by subsequent puromycin resistance selection (2 µg ml⁻¹). Small-molecule NPCs were grown in medium consisting of DMEM/F12 + Glutamax, at a ratio of 1:200 N2 supplement and 1:100 B27 supplement, penicillin-streptomycin (Life technologies), 100 mM ascorbic acid (Sigma, A4544), 3 mM CHIR99021 (Tocris, 4423) and 0.5 mM purmorphamine (Tocris, 4551) and passaged using Accutase. Generation of small-molecule NPCs from iPSCs is described in a previous publication⁵².

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

STAMP editing. For stable cell STAMP-fusion protein expression, cells were induced with 50 ng ml⁻¹ (low) or 1 µg ml⁻¹ (high) doxycycline in DMEM for 24–72 h, followed by TRIzol extraction and column purification using Direct-zol Miniprep kit (Zymo Research) in accordance with the manufacturer's protocol. Uninduced cells of the same genetic background were used as negative controls. For transient transfections, ~1 million cells were transfected with 2 µg expression construct using Fugene HD (Promega), according to the manufacturer's protocol. Upon Agilent TapeStation quantification, 500 ng 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 100 nM Torin-1 (Cell Signaling) or DMSO vehicle control alongside 1 µg ml⁻¹ doxycycline induction and collected for RNA after 72 h of incubation at 37°C.

eCLIP experiments and analysis. All STAMP-fusion (RBFox2-APOBEC1, TIA1-APOBEC1 and SLBP-APOBEC1) eCLIPs were conducted following induction or transient transfections and IP was conducted using anti-HA tag (ChIP grade;

Abcam, ab91110). eCLIP experiments were performed as previously described in a detailed standard operating procedure⁶, 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 cross-linked cells were lysed and sonicated, followed by treatment with RNase I (Thermo Fisher) to fragment RNA. Antibodies were precoupled to species-specific (anti-rabbit IgG) dynabeads (Thermo Fisher), added to lysate and incubated overnight at 4°C. Before IP washes, 2% of sample was removed to serve as the paired-input sample. For IP samples, high-salt 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 adapter was ligated with T4 RNA ligase (NEB). Ten percent 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 horseradish peroxidase-conjugated species-specific TrueBlot antibody (Rockland) and visualized with standard enhanced chemiluminescence imaging to validate successful IP. Ninety percent 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 adapter 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 adapter was ligated with T4 RNA ligase (NEB). Quantitative PCR was then used to determine the required amplification, followed by PCR with Q5 (NEB) and gel electrophoresis for size selection of the final library. Libraries were sequenced on the HiSeq 2000, 2500 or 4000 platform (Illumina). Each ENCODE eCLIP experiment consisted of IP from two independent biosamples, along with one paired size-matched input (sampled from one of the two IP lysates before IP washes). Reproducible eCLIP peaks were called using the latest release of the core pipeline (<https://github.com/yeolab/eclip/>), followed by a peak merging sub-workflow to identify reproducible peaks (https://github.com/YeoLab/merge_peaks/).

RNA-seq analysis. Bulk RNA-seq libraries were sequenced with single-end reads (100 nucleotides) and trimmed using cutadapt (v1.14.0). Trimmed reads were filtered for repeat elements using sequences obtained from RepBase (v18.05) with STAR (2.4.0i). Reads that did not map to repeats were then mapped to the hg19 assembly with STAR, sorted with samtools (v1.5) and quantified against Gencode (v19) annotations using Subread featureCounts (v1.5.3). Genes with zero counts summed across all samples were removed before performing differential expression analysis using DESeq2 (v1.26.0)⁵³.

To calculate differential expression from RNA-seq data, we used DESeq2 (v1.26.0; Supplementary Table 3), which uses a negative binomial regression model and Bayesian shrinkage estimation dispersions and fold change to estimate differentially expressed genes⁵³. Significance of logarithmic fold changes were determined by a Wald test to approximate *P* values, and genes passing an independent filtering step were adjusted for multiple testing using the Benjamini-Hochberg procedure to yield a false discovery rate (FDR). Genes with an FDR of less than 0.05 were considered statistically significant.

SAILOR calls for C-to-U edits. Resulting BAM files were each used as inputs to SAILOR (v1.1.0) to determine C > U edit sites across the hg19 assembly. Briefly, SAILOR filters potential artifacts and known SNPs (dbSNP, v147) and returns a set of candidate edit sites, evidenced by the number of C > U conversions found among aligned reads. We used an adapted Bayesian 'inverse probability model' (ref.⁵⁴) to identify high-confidence A-to-I editing sites from the RNA-seq data, where a confidence value based on the number of reads is associated with each predicted site. Sites were transformed into a broader 'window' by opening a 51-nucleotide window centered on each site.

Edit distribution, EPKM and ϵ score method details. We describe an ' ϵ score' fraction formula: $\epsilon \text{ score} = \sum_{p=1}^i \left(\frac{\sum_{q=0}^m Y_{aq}}{\sum_{c=1}^i Y_c} \right)$, where *i* represents the number of C positions *p* in a given coordinate window, with *Y_{aq}* 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 (that is, how often a C > U conversion is found) and edit potential (that is, how 'C-rich' a given region is). To find the ϵ score for a given window, we calculated the ratio between the number of (post-SAILOR-filtered) C > U read conversions to the total (post-SAILOR-filtered) coverage across every C found within the window.

To calculate EPKM values for each gene, we used cumulative edit counts (T coverage over each edit site called) as determined by SAILOR (v1.1.0). We summed region-specific (either CDS or CDS and 3' UTR, as defined by hg19 v19 Gencode annotations) edit counts for each gene and divided this number by the 'per million'

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

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

Peaks exhibiting $\log_2 FC > 2$ and $\log_{10} P > 3$ from C-terminal RBOX2-APOBEC fusion eCLIP data were shuffled within the 5' UTR, CDS and 3' UTR regions of their respective genes, over 40 permutations. These peak permutations were then expanded by 200 bp on each flank and intersected with de-noised STAMP edit clusters. The same flank expansion and intersection were conducted for the original experimentally derived eCLIP peaks. Six different versions of STAMP site 'de-noising' are reflected by the *x*-axis labels, where the decimal value reflects the confidence score used for filtering, and the 'filtered' suffix reflects application of the additional isolated site filter.

In Fig. 2c–e, RPS2–STAMP at 0, 50 and 1,000 ng doxycycline treatments compared to corresponding control–STAMP datasets were compared across gene sets taken from the Gene Expression Omnibus (GEO) under accession code GSE94460 and a previous publication³⁸. Similar comparisons were performed using normalized occupancy ratios from a previous study³⁹ (using X3 values, which closely approximate native ribosome-occupancy levels in 293T cells). In Fig. 2f, metagene profiles comparing edits (confidence level ≥ 0.5) in RPS2–STAMP, RBOX2–STAMP and control–STAMP were generated using metaPlotR (<https://github.com/olarerin/metaPlotR/>) with the highest occupancy via Ribo-seq transcripts from the GEO under accession code GSE112353, with the top quartile of expressed transcripts being used, although no expression filtering or transcript-to-gene mapping was needed as transcript-level annotations were required (Q1, $n = 4,677$). In Fig. 2g, metagene plots were generated in a similar fashion to Fig. 3g, comparing all replicates of Torin-1-treated RPS2–STAMP and vehicle-treated RPS2–STAMP. In Fig. 2h, from GEO dataset GSE94460, genes were ranked in descending order according to their replicate-averaged TPM-normalized occupancy counts. To consolidate annotations, transcripts that were found with the highest occupancy were kept. Additionally, only genes included in both of our analyses (minimally expressed protein-coding genes, TPM > 0 in either RPS2–STAMP or control–STAMP; $n = 16,128$) and the GSE112353 dataset ($n = 19,724$) were used. The remaining genes ($n = 15,485$) were divided into quartiles according to occupancy score, such that 'quartile 1' represented genes with the highest ribosome occupancy. EPKM values across CDS and 3' UTR exons within these quartiles were compared using a Wilcoxon rank-sum test to determine significance. For Extended Data Fig. 2g, similarly to Fig. 2h, Torin-1-treated control–STAMP and vehicle-treated control–STAMP were used.

Irreproducible discovery rate. IDR was used to determine reproducible edit windows between experimental replicates³³. After prefiltering SAILOR outputs with a minimum confidence score (≥ 0.5), we created 51-nucleotide windows around candidate C > U sites and calculated reproducibility scores for each window using IDR (v2.0.2). Scaled scores ($-125 \times \log_2(\text{IDR score})$) were converted to linear values and plotted, with unscaled scores ≤ 0.05 considered as reproducible sites.

RNA isolation and poly(A) selection for nanopore and PacBio sequencing. At 80% confluency in 10-cm plates, cells were washed with PBS and collected in 1 ml of TRIzol reagent (Thermo Fisher) or Direct-zol kit with DNase treatment

(Zymo Research). Total RNA was extracted following the manufacturer's protocol. Approximately 20 μg of total RNA was poly(A) selected using a poly(A) magnetic resin kit (NEB E7490L). RNA was then analyzed by high-sensitivity RNA TapeStation (Agilent, 5067-5579) to confirm poly(A) selection and RNA quality.

Direct cDNA nanopore sequencing. Next, 100 ng 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 using Oxford Nanopore PromethION flow cells (FLO-PRO002) for ~48 h. Data were base called in real time on the PromethION Guppy base callers with the high accuracy setting. Total reads (in millions) were: RBOX2 = 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 (ref. 55) with default RNA parameters. These alignments are referred to genomic and cDNA, respectively. Edits were called using Bcftools mpileup with settings '-Q 5 -d 8000 -q 1' followed by filtering each position for reference C positions on the appropriate strand. cDNA alignments were assumed to be positive stranded, and genome alignments were intersected with gene annotations to determine strand information. Sites with ambiguous strand information and/or fewer than ten reads were removed. Edit fractions were determined for sites with C-to-U mutations by the fraction (no. of mismatches)/(no. of mismatches + no. of matches). Confidence scores and SNP removal were performed via custom implementation of the SAILOR scripts. A final list of RBOX2–STAMP sites was made by subtracting all sites found in the control–STAMP with a confidence score of 0.99 or greater. Isoform-specific binding was detected by summing the number of RBOX2 unique sites and all sites identified in the control–STAMP. The top two expressing isoforms, as determined by average coverage across C positions with at least ten reads, were selected for further analysis, and isoforms comparing the largest difference in edits were compared by hand.

Direct cDNA PacBio sequencing. Technical triplicate samples containing 1 μg total RNA were extracted from HEK293T cells expressing control–STAMP and HEK293T cells expressing the RBOX2–STAMP fusions, and following 1 $\mu\text{g ml}^{-1}$ doxycycline induction for 72 h. RNA extraction was completed using Direct-zol (Zymogen). All STAMP samples were assayed for quality, and all the sample RNA integrity numbers were greater than 9. Long-read cDNA libraries were prepared according to the PacBio Iso-Seq Express protocol with 300 ng of total RNA and amplified for 13–15 cycles with the following forward and reverse primers:

Forward: 5'-GGCAATGAAGTCGCGAGGGTTG-3'

Reverse: 5'-AAGCAGTGGTATCAACGCAGAG-3'

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

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

Following sequencing, the circular consensus sequence reads for each set of technical replicates were processed using the Isoseq v3 pipeline³⁶ (<https://github.com/PacificBiosciences/IsoSeq/>) to generate full-length non-concatamer reads in fasta format. For this step, software package lima v2.0.0 was used with parameters: --isoseq and --dump-clips. In addition, isoseq3 v3.4.0 refine was used with parameters: --require-polya. Fasta files for each set of technical replicates were then pooled together and the full-length non-concatamer reads for each sample were aligned to the hg19 reference genome using minimap2 v2.17-r941 with parameters: --ax-splice, --uf, --secondary=no, --t 30. Cupcake v18.1.0 (https://github.com/Magdoll/cDNA_Cupcake/wiki/) script collapse_isoforms_by_sam.py was then run using the pooled full-length non-concatamer fasta file and aligned SAM file for each sample with parameter '--dun-merge-5-shorter' to collapse redundant isoforms. This step was completed to collapse high-quality isoforms into unique isoforms informed by genome alignment. Following this, SQANTI3 (v1.6)³⁶ (<https://github.com/ConesaLab/SQANTI3/>) script sqanti3_gc.py was used to compare the collapsed isoform results from Cupcake to the Gencode hg19 (v19) annotation to characterize the collapsed isoforms.

Edit fractions were quantified for each sample using the SAILOR computational tool without filtering reads for RBOX2–APOBEC1 and APOBEC1–control samples. Edited positions with a confidence score of greater than or equal to 0.99 were then used to elucidate motifs using HOMER tool findMotifsGenome.pl (v4.9.1)³⁷

A custom script was generated to quantify the percentage of edited reads in the 3' exonal region for each sample. Only previously annotated genes were considered, using the Gencode hg19 (v19) annotation as the reference. For each gene, the isoforms associated with the gene were first determined based on assignment by the SQANTI3 isoform classification pipeline. Only genes with two

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

Single-cell RNA-seq. For the scRNA-seq of transduced cells, following 72 h of doxycycline treatment ($1 \mu\text{g ml}^{-1}$), cells were trypsinized (TrypLE, Invitrogen), counted and resuspended at a density of 1,000 cells per 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), per the manufacturer's protocol. In total, 16,000 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' adapter and sample index attachment as indicated in the manufacturer's protocol for 3' expression capture. The 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 oligonucleotide primers and finally indexed. Agilent High Sensitivity D5000 ScreenTape Assay (Agilent Technologies) was performed for quality control of the libraries. The 3' poly(A) and feature libraries were sequenced on an Illumina NovaSeq 6000. Following 3' poly(A) read demultiplexing, alignment to the hg19 and custom hg19 and lentiviral-genes transcriptomes and unique molecular identifier -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 package³⁸. Matrices for all samples were concatenated when necessary, and all genes that were not detected in at least 0.1% of single cells were discarded. Cells with fewer than 1,000 or more than 7,000 expressed genes, as well as cells with more than 50,000 unique transcripts or 20% mitochondrial expressed genes, were removed from the analysis. The only exception for these filters was for the NPC:HEK293T samples for which only cells with over 25% mitochondrial genes were filtered out and cell doublets were removed with Scrublet. Transcripts per cell were normalized to 10,000, a unit was added and scores were logarithmized ($\ln(\text{TPM} + 1)$) and scaled to unit variance (z -scored). The top 2,000 variable genes were identified with the `filter_genes_dispersion` function, `flavor = 'cell_ranger'`. Principal component analysis was carried out, and the top 40 principal components were retained. With these principal components, neighborhood graphs were computed with ten neighbors and standard parameters with the `pp.neighbors` function. Single-cell edits were called by first computing the MD tag from Cell Ranger outputs (`possorted_genome_bam.bam`) using Samtools 'calmd' and splitting every read according to their cell barcode. SAILOR, ϵ score, region EPKM-score and motif analyses were run for each cell (or the aggregate of reads for all cell barcodes within defined Louvain clusters) in a similar fashion to bulk RNA-seq. 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 Scanpy (v1.4.4)³⁸. Matrices for all samples were concatenated, and all genes that were not edited in at least two 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-STAMP, NPC:HEK293T-RBFOX2-STAMP, HEK293T-control-STAMP, HEK293T-RBFOX2:TIA1-STAMP, HEK293T-RPS2-STAMP and RPS2-cluster cells, respectively. Cells with fewer than ten edited genes were removed from the analysis. EPKM or ϵ scores for each cell were normalized to 10,000, a unit was added and scores were logarithmized ($\ln(\text{TPM} + 1)$) and scaled to unit variance (z -scored). Principal component analysis was carried out, and the top 40 principal components were retained. With these principal components, neighborhood graphs were computed with ten 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 to define RBP-specific clusters. Single-cell and mean ϵ scores per the sample heat maps were generated with the `pl.heatmap` and `pl.matrixplot` functions, respectively. Differentially edited genes were determined for each set of Louvain (or modified) clusters with the `tl.rank_gene_groups` function (`method = 'wilcoxon'`).

In Fig. 4g, edits were called from groups of randomly selected RBFOX2-capture sequence barcodes ($n = 1, \dots, 49, 100, 200, 300, 400, 500, 600,$

700, 800 and 844 cells) and processed using the SAILOR pipeline. To discover whether or not sites were globally enriched for known binding motifs, we recalculated the confidence score using the same ' ϵ score' (number of $C > U$ read conversions over the total coverage across all C positions within a window) across all 51-nucleotide windows surrounding each candidate edit site and filtered these windows using various scores (0.99 and 0.999). We performed de novo motif finding using HOMER (v4.9.1) using these filtered windows and a shuffled background for each UTR, CDS, intron and total genic region (`findMotifs.pl foreground.fa fasta outloc --nofacts --p 4 --rna --S 20 --len 6 --noconvert --nogo --fasta background.fa`), resulting in a set of fasta sequences corresponding to each 51-nucleotide edit window, as well as a corresponding random background. This was repeated ten 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 ((U)GCAUG) was then used as a pivot, and significance was recalculated 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`).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw and assembled sequencing data from this study have been deposited in NCBI's Gene Expression Omnibus (GEO) under accession code [GSE155729](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155729). Processed edit coordinates are available in Supplementary Tables 1, 2 and 4. Differential edit and gene expression data are available in Supplementary Tables 3 and 5–9. Published ribosome profiling data used in this study are deposited in the GEO under accession code [GSE94460](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94460) and polysome sequencing data are deposited in the GEO under accession code [GSE109423](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109423).

Code availability

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

References

- Li, Y. et al. A comprehensive library of familial human amyotrophic lateral sclerosis induced pluripotent stem cells. *PLoS ONE* **10**, e0118266 (2015).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Gordon, S. P. et al. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS ONE* **10**, e0132628 (2015).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

Acknowledgements

We thank G. Viero at the Institute of Biophysics, CNR Unit at Trento for helpful advice concerning analysis of Ribo-seq data. We thank T. Michael at the J. Craig Venter Institute for use of the Oxford Nanopore PromethION system. We are grateful to K. D. Meyer for the gift of the YTH-APOBEC1 construct. We thank J. Rothstein at Johns Hopkins University School of Medicine for the gift of the NPC cell line. We are grateful to the La Jolla Institute's Immunology Sequencing Core and the IGM Genomics Center, University of California San Diego, for use of the 10x Chromium and Illumina sequencing platforms. This work was partially supported by the National Institutes of Health (NIH; HG004659 and HG009889 to G.W.Y.). I.A.C. is a San Diego IRACDA Fellow supported by an NIH/NIGMS award (K12GM068524). R.J.M. was supported in part by an institutional award to the UCSD Genetics Training Program from the National Institute for General Medical Sciences, (T32GM008666) and a Ruth L. Kirschstein National Research Service award (F31NS111859). K.W.B. is a University of California President's Postdoctoral Fellow supported by an NIH/NINDS Career Transition Award (K22NS112678).

Author contributions

Conceptualization: K.W.B. and G.W.Y.; methodology: K.W.B. and G.W.Y.; investigation: K.W.B., I.A.C., K.D.D., A.A.M., D.A.L. and R.J.M.; formal analysis: E.K., P.J., K.W.B.,

B.A.Y., I.A.C., D.A.L. and R.J.M.; writing of original draft: K.W.B. and G.W.Y.; writing of review and editing: G.W.Y., K.W.B., I.A.C., B.A.Y., D.A.L. and R.J.M.; funding acquisition: G.W.Y.; supervision: G.W.Y.

Competing interests

G.W.Y. is a cofounder, a member of the Board of Directors, on the Scientific Advisory Board, an equity holder and a paid consultant for Locanabio and Eclipse BioInnovations. G.W.Y. is a visiting professor at the National University of Singapore. G.W.Y.'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 interests.

Additional information

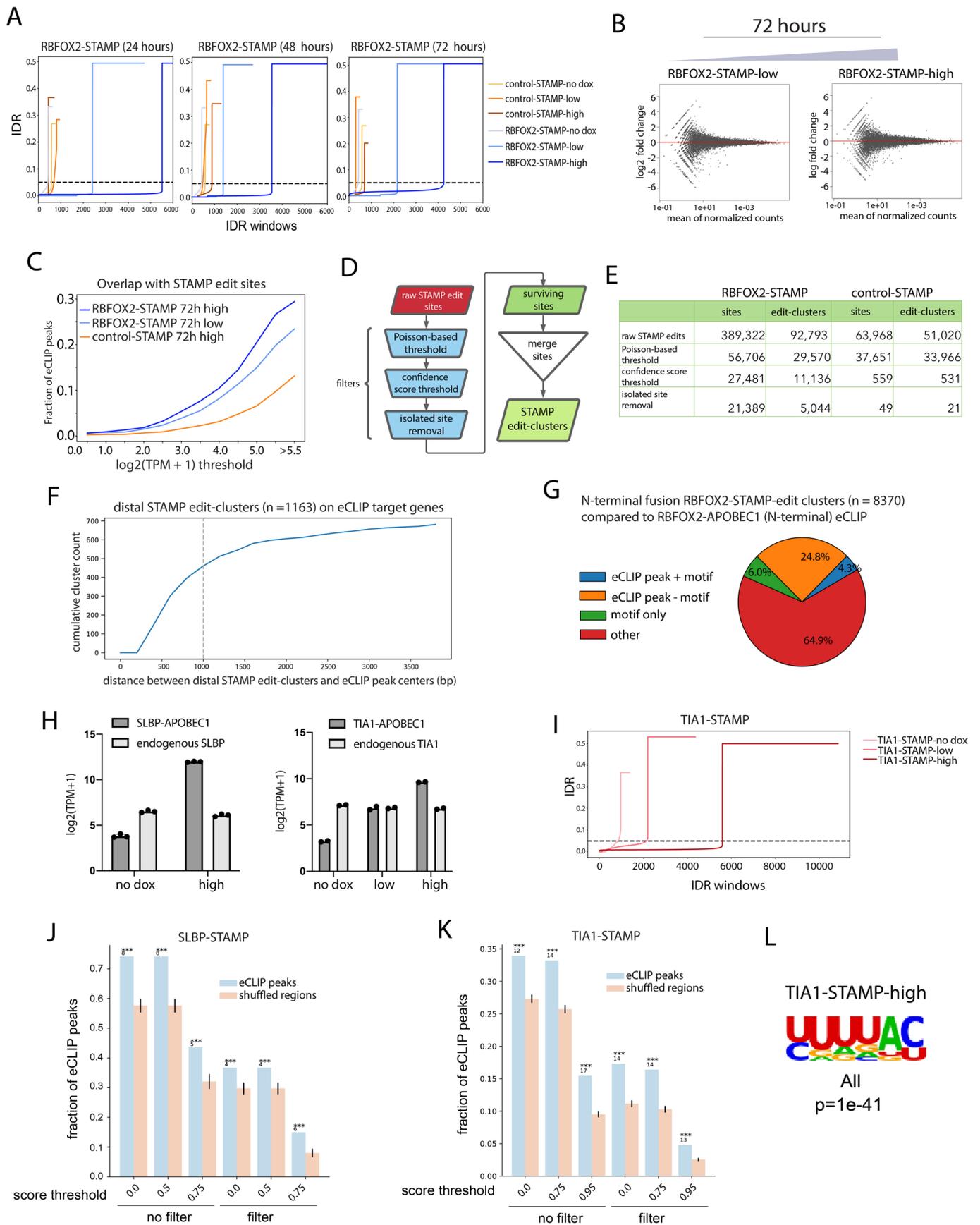
Extended data is available for this paper at <https://doi.org/10.1038/s41592-021-01128-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-021-01128-0>.

Correspondence and requests for materials should be addressed to G.W.Y.

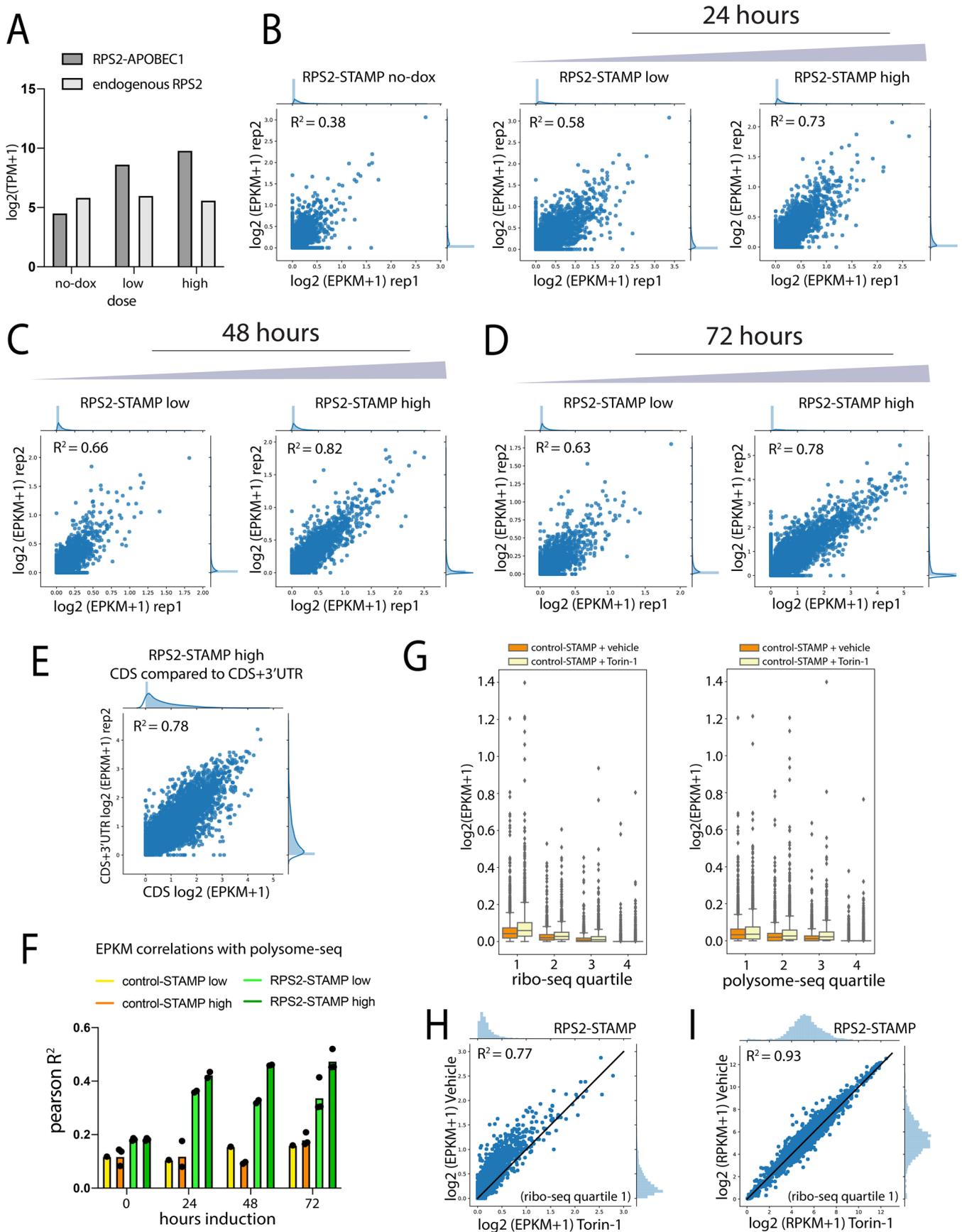
Peer review information *Nature Methods* thanks Alfredo Castello and the other, anonymous, reviewers for their contribution to the peer review of this work. Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.



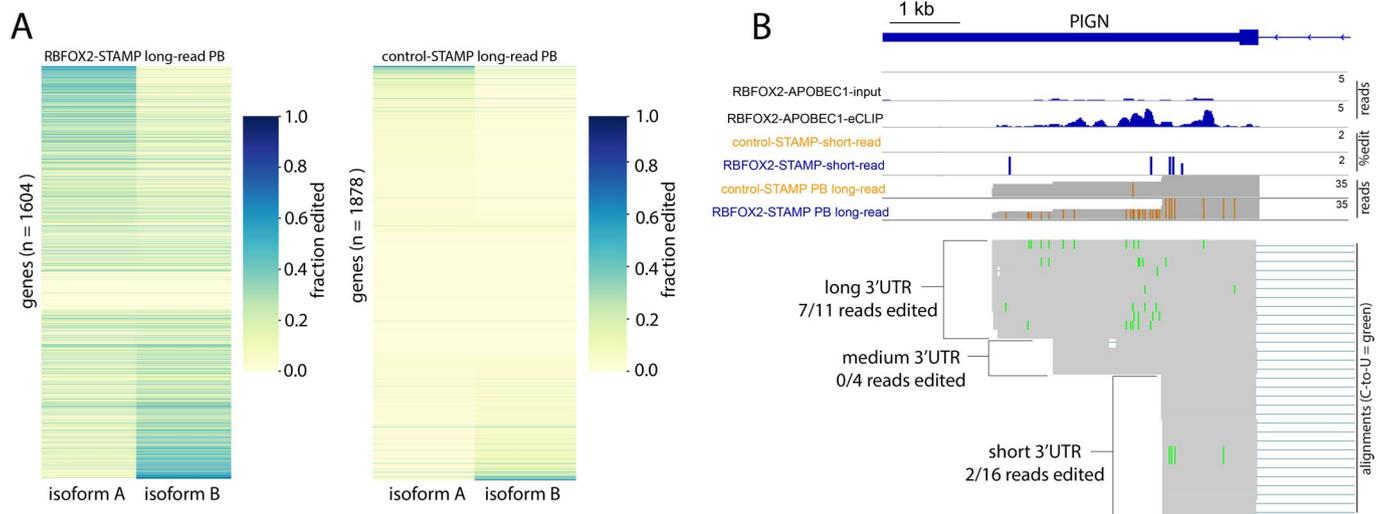
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | RBP-STAMP reproducibility and concordance with eCLIP, related to Figure 1. a, Irreproducible Discovery Rate (IDR) analysis comparing ≥ 0.5 confidence edit windows for increasing levels of RBFOX2-STAMP at 24, 48 and 72 hours. **b**, Differential expression (DEseq2) analysis of RBFOX2-STAMP for increasing levels of RBFOX2-STAMP at 72 hours. **c**, Fraction of RBFOX2-APOBEC1 eCLIP peaks overlapping low and high induction RBFOX2-STAMP edit sites at increasing expression (TPM) thresholds. **d**, STAMP edit-site filtering and cluster-calling workflow. **e**, Number of control- and RBFOX2-STAMP edit sites and clusters retained after each filtering step in D. **f**, Cumulative distance measurement from RBFOX2-STAMP distal edit-clusters to eCLIP peaks on targets genes. **g**, Pie chart showing the proportion of N-terminally fused RBFOX2-APOBEC1 STAMP edit-clusters overlapping with either 1) RBFOX2-APOBEC1 N-terminal fusion high-confidence eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over input) containing the conserved RBFOX2 binding motif (GCAUG), 2) equally stringent eCLIP peaks not containing the conserved motif, 3) the conserved motif falling outside of eCLIP peaks, or 4) neither eCLIP peaks nor conserved motifs. **h**, Quantification of expression from no dox (0ng/ml) low (50ng/ml) or high (1 μ g/ml) doxycycline induction of SLBP-APOBEC1 and TIA1-APOBEC1 fusions compared to endogenous expression. **i**, Irreproducible Discovery Rate (IDR) analysis comparing $0.5 \geq$ confidence level edit windows for increasing levels of TIA1-STAMP at 72 hours. **j**, Fraction of SLBP eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over size-matched input, reproducible by IDR) with SLBP-STAMP edit-clusters, compared to size-matched shuffled regions, calculated at different edit site confidence levels before and after site filtering (see Materials and Methods for filtering procedure). Numbers atop bars are Z-scores computed comparing observed with the distribution from random shuffles. *** denotes statistical significance at $p = 0$, one-sided exact permutation test. **k**, Fraction of TIA1-APOBEC1 eCLIP peaks ($\log_2fc > 2$ and $-\log_{10}p > 3$ over size-matched input) with TIA1-STAMP edit-clusters, compared to size-matched shuffled regions, calculated at different edit site confidence levels before and after site filtering (see Materials and Methods for filtering procedure). Numbers atop bars are Z-scores computed comparing observed with the distribution from random shuffles. *** denotes statistical significance at $p = 0$, one-sided exact permutation test. **l**, Motif enrichment using HOMER and shuffled background on TIA1-STAMP edit-clusters.

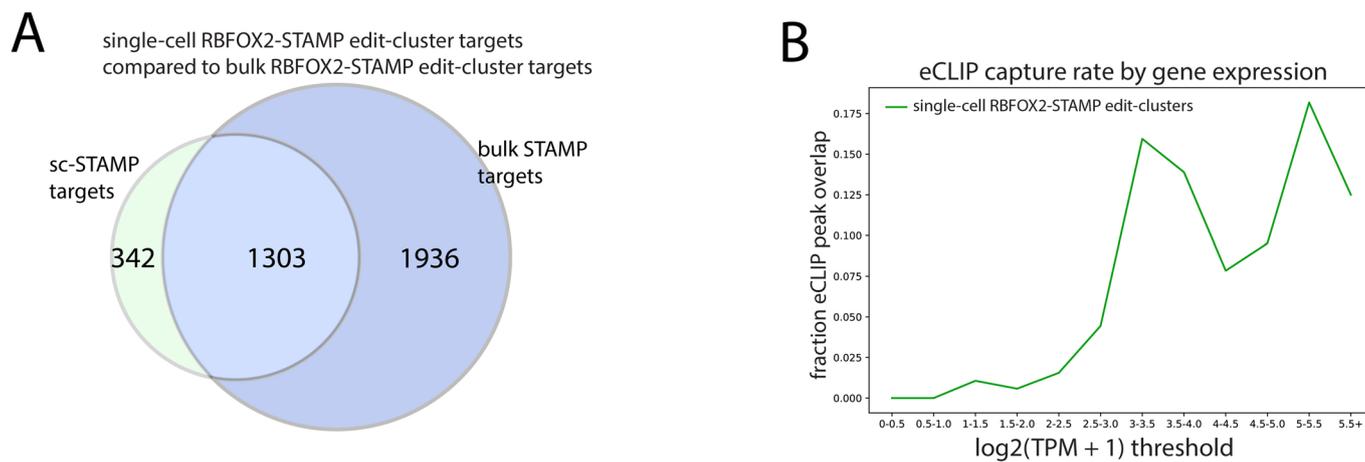


Extended Data Fig. 2 | See next page for caption.

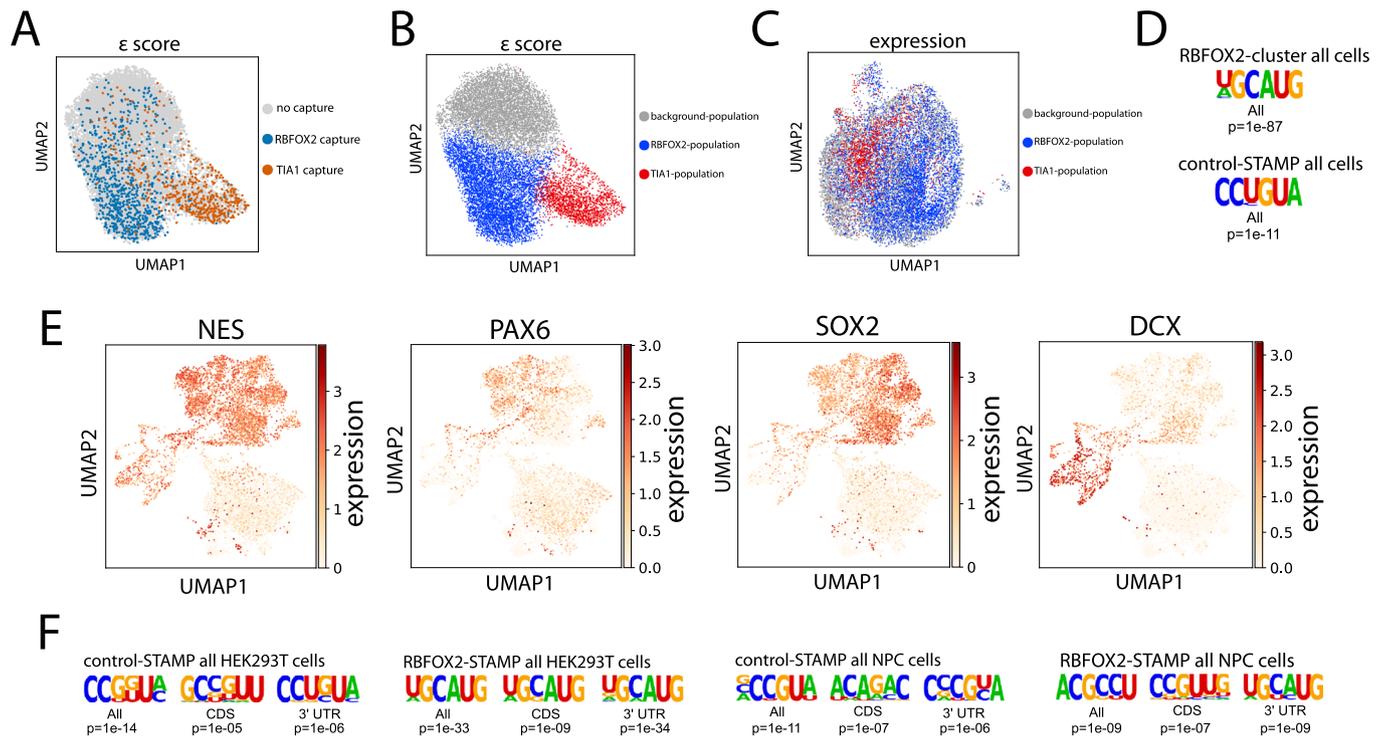
Extended Data Fig. 2 | Ribo-STAMP reproducibility and response to mTOR pathway perturbations, related to Figure 2. **a**, Quantification of expression from no dox (0ng/ml) low (50ng/ml) or high (1µg/ml) doxycycline induction of RPS2-APOBEC1 fusion compared to endogenous expression. **b–d**, Scatterplot comparisons of CDS+3'UTR EPKM values from RPS2-STAMP replicate experiments showing high, dose-dependent correlation at 24 (B), 48 (C) and 72 hours (D). **e**, Scatterplot comparison of CDS EPKM values with CDS+3'UTR EPKM values for RPS2-STAMP. **f**, Pearson R^2 values for low and high induction control- or RPS2-STAMP EPKM compared to poly-ribosome-enriched polysome-seq RPKM. **g**, Comparison of EPKM from vehicle treated 72-hour high-induction control-STAMP compared to Torin-1 treated 72-hour high-induction control-STAMP showing no significant signal reduction for top ribosome occupied quartile genes containing Torin-1 sensitive TOP genes as detected by ribo-seq (Q1 $p = 1.0$, $n = 3589$ genes, Wilcoxon rank-sum one-sided) and polysome profiling (Q1 $p = 1.0$, $n = 3589$ genes, Wilcoxon rank-sum one-sided). **h**, Scatterplot comparison of CDS+3'UTR EPKM values on ribo-seq top quartile genes ($n = 3589$) for Torin-1 treated and vehicle treated RPS2-STAMP 72-hour high (1µg/ml) doxycycline inductions as in Figure 2H. **i**, Scatterplot comparison of CDS+3'UTR RPKM values on ribo-seq quartile-1 genes ($n = 3589$) for Torin-1 treated and vehicle treated RPS2-STAMP 72-hour high (1µg/ml) doxycycline inductions.



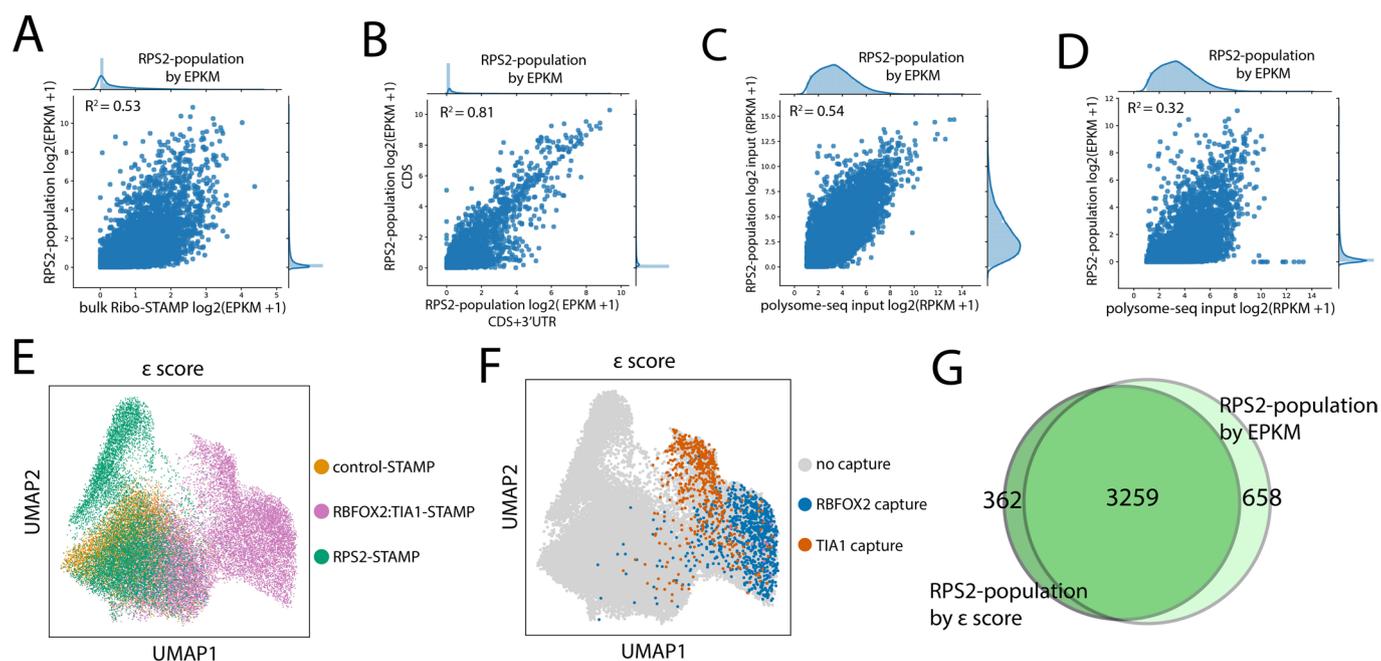
Extended Data Fig. 3 | Long-read STAMP reveals isoform specific binding profiles, related to Figure 3. a, Heatmap of control- and RBFOX2-STAMP edit fractions calculated from the final exon of all detected primary and secondary alternative polyadenylation (APA) isoforms meeting coverage criteria (see materials and methods). **b**, IGV tracks showing RBFOX2-APOBEC1 eCLIP peaks, control- and RBFOX2-STAMP short-read edit clusters, compared to control- and RBFOX2-STAMP long-read (PB) alignments on long, middle and short APA isoforms of the target gene *PIGN*, with green colored C-to-U conversions on different isoforms.



Extended Data Fig. 4 | Comparison of bulk STAMP to single-cell STAMP, related to Figure 4. **a**, Overlap between single-cell and bulk RBFox2-STAMP target genes containing edit-clusters. **b**, Fraction of RBFox2-APOBEC1 eCLIP peaks overlapping low and high induction single-cell RBFox2-STAMP edit-clusters at increasing expression (TPM) thresholds.



Extended Data Fig. 5 | Single-cell RBP-RNA interaction detection by STAMP for multiple RBPs and in multiple cell types, related to figure 5. a, UMAP plot using ϵ score from RBFOX2-STAMP and TIA1-STAMP mixture with capture sequence RBFOX2-STAMP (blue, $n = 844$) and TIA1-STAMP cells (red, $n = 527$) highlighted. **b**, UMAP plot as in **a** color-coded by Louvain clustering into RBFOX2-cluster (blue), and TIA1-cluster (red), or background-cluster (gray) populations. **c**, UMAP plot of gene expression for ϵ score Louvain clusters defined in **b**. **d**, Motif enrichment using HOMER from ≥ 0.99 confidence edits from combined RBFOX2-cluster and control-STAMP cells. **e**, UMAP plot showing expression of neural precursor cell markers *NES*, *PAX6*, *SOX2* and *DCX*. **f**, Motif enrichment using HOMER from ≥ 0.99 confidence edits from combined control- and RBFOX2-STAMP HEK293T and NPC cells.



Extended Data Fig. 6 | Single Ribo-STAMP detects ribosome occupancy from individual cells, related to Figure 6. a, Genome-wide comparison of CDS+3'UTR EPKM values for bulk and single-cell EPKM-derived RPS2-population. **b**, Comparison of EPKM-derived RPS2-population CDS and CDS+3'UTR EPKM values. **c**, Comparison of EPKM-derived RPS2-population total mRNA RPKM values with total mRNA RPKM values from polysome-seq input. **d**, Comparison of EPKM-derived RPS2-population CDS+3'UTR EPKM values with total mRNA RPKM values from polysome-seq input. **e**, UMAP analysis of ϵ score from merged 72-hour high-induction RPS2-STAMP (green), control-STAMP (orange) and mixed-cell RBFOX2:TIA1-STAMP (purple) single-cell experiments. **f**, UMAP plot as in E with only capture sequence RBFOX2-STAMP (blue, $n = 844$) and TIA1-STAMP cells (red, $n = 527$) highlighted. **d**, Individual cell barcode overlap for EPKM-derived and ϵ score-derived RPS2-populations.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
 - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
 - The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
 - A description of all covariates tested
 - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
 - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
 - For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
 - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
 - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
 - Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	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).
-----------------	--

Data analysis	https://github.com/YeoLab/SAILOR
---------------	---

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data are available at GSE155729, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155729>

All plasmids are available from the authors upon request.

(bulk RNAseq) <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155649> (reviewer/private access token: cbmfaeyixtsxnwf)
 (Nanopore) <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155708> (reviewer/private access token: iridycagpzalze)
 (scRNA) <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155726> (reviewer/private access token: ehikhwqgndszyt)
 (eCLIP) <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155728> (reviewer/private access token: szwtomsspnwppil)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. For EPKM scatterplot and quartile comparisons, sample size is simply gene number within each comparison or quartile with detectable signal, which is an unbiased representation of signal from all available genes. The number of genes used to analyze and benchmark editing profiles was a byproduct of editing efficiency and gene expression. Genes below a certain expression threshold were excluded from analyses to avoid potential noise
Data exclusions	There are no data exclusions
Replication	All replicates are included. Each standard STAMP experiment was conducted with at least one technical replicate, and a maximum of 3 replicates.
Randomization	Allocation of experimental groups was not random — each STAMP fusion experiment was subjected to the same APOBEC-fusion and dox-induction procedures and timing. All were compared to a non-RBP APOBEC-only control.
Blinding	Blinding was not relevant to our experiment given that each RBP had previously known binding profiles that were needed to benchmark STAMP, and ribo-STAMP metrics such as EPKM were unrelated to metrics used for other RBPs. Thus each experiment required a priori knowledge of which RBP was being probed in order to properly assess STAMP efficacy.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Anti-HA tag antibody - ChIP Grade (Abcam, ab91110)
Validation	Tested applications Suitable for: ChIP/Chip, IP, ELISA, WB, ICC/IF, ICC, Flow Cyt, ChIP, as tested by Abcam company.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

Lenti-X HEK 293T Cell Line (Female) Clontech Cat# 632180
HEK293T control-STAMP (APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T RBFOX2-STAMP (RBFOX2-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T RBFOX2-Capture-1-STAMP (RBFOX2-Capture-1-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T SLBP-STAMP (SLBP-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T TIA1-STAMP (TIA1-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T TIA1-Capture-1-STAMP (TIA1-Capture-1-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T RPS2-STAMP (RPS2-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T RPS2-Capture-1-STAMP (RPS2-Capture-1-APOBEC1-HA-P2A-mRuby) This paper N/A
HEK293T RPS3-STAMP (RPS3-APOBEC1-HA-P2A-mRuby) This paper N/A
iPSCs derived smNPC
Gift from Jeff Rothstein N/A

Authentication

None of the cell lines were authenticated

Mycoplasma contamination

All cell lines tested negative for mycoplasma contamination

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified lines were used in this study