

Long-read Ribo-STAMP simultaneously measures transcription and translation with isoform resolution

Pratibha Jagannatha,^{1,2,3,4} Alexandra T. Tankka,^{1,2,3} Daniel A. Lorenz,⁵ Tao Yu,^{1,2,3} Brian A. Yee,^{1,2,3} Kristopher W. Brannan,^{1,2,3,7} Cathy J. Zhou,^{1,2,3} Jason G. Underwood,⁶ and Gene W. Yeo^{1,2,3,4,5}

¹Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California 92093, USA; ²Sanford Stem Cell Institution Innovation Center and Stem Cell Program, University of California San Diego, La Jolla, California 92037, USA;

³Institute for Genomic Medicine, University of California San Diego, La Jolla, California 92093, USA; ⁴Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, California 92093, USA; ⁵Sanford Laboratories for Innovative Medicine, La Jolla, California 92121, USA; ⁶PacBio, Menlo Park, California 94025, USA

Transcription and translation are intertwined processes in which mRNA isoforms are crucial intermediaries. However, methodological limitations in analyzing translation at the mRNA isoform level have left gaps in our understanding of critical biological processes. To address these gaps, we developed an integrated computational and experimental framework called long-read Ribo-STAMP (LR-Ribo-STAMP) that capitalizes on advancements in long-read sequencing and RNA-base editing–mediated technologies to simultaneously profile translation and transcription at both the gene and mRNA isoform levels. We also developed the EditsC metric to quantify editing and leverage the single-molecule, full-length transcript information provided by long-read sequencing. Here, we report concordance between gene-level translation profiles obtained with long-read and short-read Ribo-STAMP. We show that LR-Ribo-STAMP successfully profiles translation of mRNA isoforms and links regulatory features, such as upstream open reading frames (uORFs), to translation measurements. We apply LR-Ribo-STAMP to discovering translational differences at both the gene and isoform levels in a triple-negative breast cancer cell line under normoxia and hypoxia and find that LR-Ribo-STAMP effectively delineates orthogonal transcriptional and translation shifts between conditions. We also discover regulatory elements that distinguish translational differences at the isoform level. We highlight *GRK6*, in which hypoxia is observed to increase expression and translation of a shorter mRNA isoform, giving rise to a truncated protein without the AGC Kinase domain. Overall, LR-Ribo-STAMP is an important advance in our repertoire of methods that measures mRNA translation with isoform sensitivity.

[Supplemental material is available for this article.]

Post-transcriptional processes such as alternative splicing (AS) and polyadenylation result in mRNA isoforms that differ in their coding and noncoding sequences, resulting in a diversity of abundance and localization of protein isoforms (Kornblihtt et al. 2013; Mitschka and Mayr 2022). Long-read sequencing has enabled the examination of full-length transcriptomes at single-molecule resolution (Foord et al. 2023). However, commonly used methods that probe mRNA translation, such as ribosome footprinting (Ribo-seq) (Ingolia et al. 2009) and polysome profiling, are incompatible with long-read sequencing owing to fragmentation and high input material requirements, respectively. Both approaches also require the separate generation of RNA-seq libraries in parallel with ribosome-protected fragments (in Ribo-seq) or polysome-fractionated mRNAs (in polysome profiling) to compute translation efficiency (TE). Although computational attempts have been proposed to better understand the components that affect translation (Gunawardana and Niranjana 2013; Cui et al. 2019; Li et al. 2019; Quattrone and Dassi 2019;

Reixachs-Solé et al. 2020), these approaches still reflect the limitations of the methods above. Transcript isoforms in polysome sequencing (TriP-seq) (Floor and Doudna 2016) and fractionation and high-throughput RNA sequencing (Frac-seq) (Sterne-Weiler et al. 2013) have been used to quantify isoform-level translation by coupling polysome association with short-read sequencing. However, using short reads to quantify isoforms is challenging owing to ambiguities in mapping reads to specific isoforms, particularly in regions with high sequence similarity. Consequently, most studies of mRNA translation have focused only on gene level differences.

Recently developed technologies that fuse RNA base editors (rBEs) to full-length proteins have been used to effectively identify regions of protein–RNA interactions with compatibility with long-read sequencing and minimal input (Brannan et al. 2021; Flamand et al. 2022; Lin et al. 2023). The surveying targets by APOBEC1-mediated profiling (STAMP) methodology first introduced the Ribo-STAMP concept to simultaneously measure gene-level mRNA levels and ribosome association by fusing ribosomal subunit proteins to cytosine deaminase enzyme APOBEC1 (Brannan et al. 2021). Ribo-STAMP was combined with short-read sequencing and

⁷Present address: Center for RNA Therapeutics, Department of Cardiovascular Sciences, Houston Methodist Research Institute, Houston, TX 77030, USA

Corresponding author: geneyeo@ucsd.edu

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.279176.124>. Freely available online through the *Genome Research* Open Access option.

© 2024 Jagannatha et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

single-cell capture approaches to successfully measure translomes at the gene level.

In this study, we develop an integrated computational and experimental framework to couple Ribo-STAMP with long-read sequencing to provide simultaneous transcription and translation measurement at mRNA isoform resolution in a single preparative step. We demonstrate that we can profile transcription and translation in a scalable manner at both the gene and isoform levels in unperturbed cells. In addition, we show that long-read Ribo-STAMP (LR-Ribo-STAMP) is an effective tool for distinguishing transcriptional and translational changes and regulatory rules, such as in a cellular model of triple-negative breast cancer (MDA-MB-231) under hypoxic conditions.

Results

LR-Ribo-STAMP experimental and computational overview

LR-Ribo-STAMP combines Ribo-STAMP technology with long-read sequencing platforms to enable translation profiling and quantification at both the gene and isoform levels. Here, ribosomal protein S2 (RPS2) is fused to APOBEC1, an RNA editing enzyme that catalyzes C-to-U editing on RNA transcripts, to enable simultaneous measurements of mRNA translation and mRNA levels (Fig. 1A).

To identify and quantify Ribo-STAMP editing and infer translation levels at gene and mRNA isoform levels, we developed a sequencing platform-agnostic computational pipeline involving long-read alignment and filtering, transcript quantification and read assignment to isoforms (Methods), identification and filtering of edited sites, and calculation of the ratio of edited to total cytosines across a gene's or isoform's exons (EditsC) (Fig. 1B). We developed the EditsC metric to capitalize on the single-molecule, full-length transcript information provided by long-read sequencing, enabling the precise identification of all editable cytosines within individual RNA molecules and their associated isoforms. This is in contrast to short-read sequencing, which only provides a generalized view of editing events aggregated at the gene level, without the ability to differentiate between isoforms. Although users can utilize the pipeline for unannotated isoforms, this report focuses on annotated isoforms. Although it can identify edits across transcripts, we concentrate on those within the coding sequence (CDS), as ribosomes predominantly interact with transcripts in this region.

A crucial aspect of the computational pipeline is distinguishing signal from background, in which the background can result from single-nucleotide polymorphisms (SNPs) and native RNA editing. Existing tools for RNA editing detection in long-read data enforce strict thresholding on editing fractions—the ratio of edited to

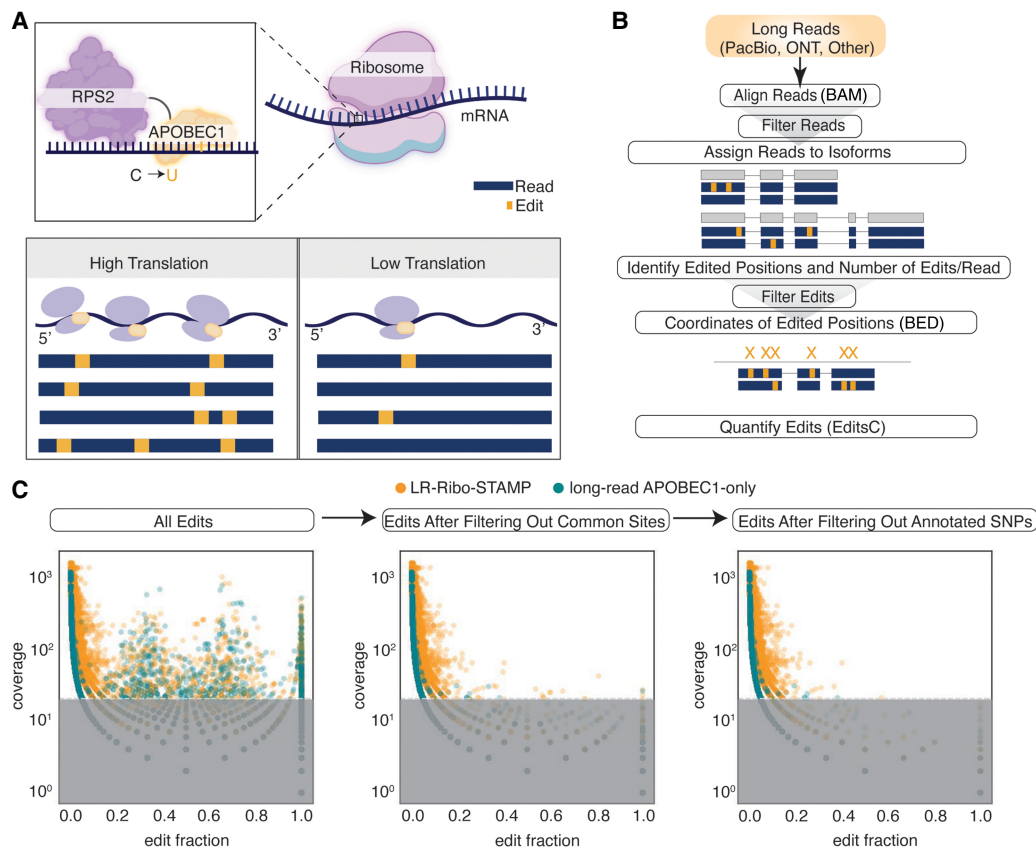


Figure 1. Experimental and computational methods for long-read Ribo-STAMP (LR-Ribo-STAMP). (A) Overview of the LR-Ribo-STAMP experimental system. RPS2 is fused to APOBEC1 to induce cytosine-to-uracil nucleotide edits proximal to ribosome–RNA interaction sites. More edits indicate higher translation, and fewer edits indicate lower translation. (B) Overview of the LR-Ribo-STAMP computational pipeline. The input is unaligned long reads, which undergo alignment, read filtering, edit detection, edit filtering, and edit quantification. Edited sites are output as a BED file. (C) Edit filtering. An outline of edit filtering and delineation of LR-Ribo-STAMP (gold) from long-read APOBEC1-only (green) signal through filtering common sites and annotated SNPs represented as the relationship between edit fraction (edited reads/total number of reads) versus coverage at an edited site. Edited sites in the gray portion indicate sites having fewer than 20 reads filtered out.

total reads at a position (Liu et al. 2023b). To minimize signal loss, we adopted an alternative filtering method to accommodate the variable nature of Ribo-STAMP editing and the challenge of estimating edit fractions at low coverage sites for which data are sparse. Our approach leverages a read threshold that only considers positions with more than 20 reads (Supplemental Fig. 1A) to ensure reliable edit detection and eliminates edited sites found across sample types and replicates, reasoning that they would be unlikely owing to the transient nature of ribosome association with mRNA transcripts. Testing this hypothesis with Pacific Biosciences (PacBio) long-read data obtained from HEK293T cells expressing Ribo-STAMP or APOBEC1-only constructs, we observed effective removal of overlapping signal found in Ribo-STAMP and APOBEC1-only controls, with most of the removed edited sites corresponding to annotated SNPs (Fig. 1C; Supplemental Fig. 1B). Although our approach does not explicitly filter based on the fraction of reads edited at a site, the edit fraction is visualized to illustrate that sites with a higher edit fraction—in line with SNP sites—are more likely to be removed by our methodology. Users can use an annotated SNP database to further filter edited sites. To ensure that the EditsC metric calculated after edit filtering is not biased by the number of cytosines in an isoform, we looked at the correlation between the average EditsC calculated at the isoform level for replicates of LR-Ribo-STAMP and long-read APOBEC1-only and the number of cytosines in an isoform and found no correlation (Supplemental Fig. 1C).

LR-Ribo-STAMP profiles gene-level translation in unperturbed cells

Ribo-STAMP was initially developed and extensively benchmarked for short-read sequencing (on the Illumina HiSeq platform) to measure gene-level translation (Brannan et al. 2021). To determine if LR-Ribo-STAMP could similarly quantify translation, we generated separate HEK293T cell lines stably integrated with doxycycline (dox)-inducible Ribo-STAMP and APOBEC1-only expression vectors. Editing found in Ribo-STAMP samples represents ribosome association with transcripts, and those in APOBEC1-only samples represent the editing profile of the enzyme alone. We generated PacBio-compatible long-read cDNA sequencing libraries from these cell lines induced with dox for 72 h and sequenced three replicates each for LR-Ribo-STAMP and long-read APOBEC1-only. A comparative analysis using long-read data revealed similar numbers of mapped reads across samples, but APOBEC1-only samples showed less editing than did LR-Ribo-STAMP, leading to larger average EditsCs in LR-Ribo-STAMP across replicates (Supplemental Fig. 1D–G).

After filtering for genes having at least 20 reads for all replicates in LR-Ribo-STAMP and long-read APOBEC1-only samples, we quantified editing for 428 genes that contained adequate read counts and quantifiable editing and found enriched editing in LR-Ribo-STAMP over long-read APOBEC1-only samples (Fig. 2A). Because the cells are unperturbed in both sample types, we expected the difference in signal to stem primarily from differences in editing levels rather than differential gene expression. Using principal component analysis (PCA), we observed distinct clustering of LR-Ribo-STAMP and long-read APOBEC1-only samples by EditsC as opposed to reads per kilobase per million (RPKM) calculated from the long reads, confirming editing levels to be the causative factor (Fig. 2B).

To assess the agreement between Ribo-STAMP data generated with long-read and short-read sequencing, we compared our

gene-level quantifications from LR-Ribo-STAMP (measured as EditsC) against short-read Ribo-STAMP (measured as edits per kilobase per million [EPKM]) quantified in our group's earlier publication (Brannan et al. 2021). Pearson's correlation was used to assess the agreement, and we confirmed statistically significant and positive concordance between the two data sets ($R = 4.09 \times 10^{-1}$, $P = 2.36 \times 10^{-48}$) (Fig. 2C). Furthermore, we explored the possibility that LR-Ribo-STAMP's accumulation of ribosome association information over time might correlate to protein production levels, a connection not previously studied with short-read Ribo-STAMP data. Correlation between the gene-level quantification from long-read data and previously published mass spectrometry data for steady-state HEK293T cells (Hegazi et al. 2022) showed a stronger positive correlation in LR-Ribo-STAMP samples ($R = 4.03 \times 10^{-1}$, $P = 7.53 \times 10^{-14}$) relative to long-read APOBEC1-only, suggesting that LR-Ribo-STAMP is a useful proxy for protein abundance (Fig. 2D; Supplemental Fig. 1H).

In addition to background signals stemming from SNPs and native RNA editing, RNA-base editing-mediated technologies are susceptible to spurious editing and biases of the rBEs they fuse to (Medina-Munoz et al. 2024). Although this background may be less of an issue when comparing editing of the same gene or isoform across conditions in which relative changes are more discernable, they become more confounding when comparing editing between different genes and isoforms in unperturbed conditions. To address this background, we performed linear regression using EditsC calculated for LR-Ribo-STAMP and long-read APOBEC1-only samples to identify genes with the highest signal-to-background ratio. We found 141 genes after thresholding with a standard deviation threshold of 1 for residual values (Fig. 2E). Gene Ontology analysis showed that these genes had a higher association with RNA processing, translation, and cell cycle processes, as may be expected of proliferating cells in unperturbed conditions, unlike the top genes identified only by LR-Ribo-STAMP EditsC ranking (Fig. 2F; Supplemental Fig. 1I). These results suggest that the linear model effectively identifies genes with strong signals. Overall, LR-Ribo-STAMP demonstrates substantial positive concordance with previously benchmarked short-read Ribo-STAMP data and can effectively profile gene-level translation quantification and, potentially, protein production in unperturbed cells.

LR-Ribo-STAMP profiles mRNA isoform translation in unperturbed cells

As potentially the first technology for concurrent transcription and translation measurements, we assessed if LR-Ribo-STAMP can indeed profile mRNA isoform translation at isoform-level resolution. To do so, we focused on annotated isoforms with at least 20 long reads across all replicates regardless of isoform length and called edits (Supplemental Table 1). We quantified editing for 405 annotated mRNA isoforms. We observed higher levels of editing in LR-Ribo-STAMP than in long-read APOBEC1-only samples (Supplemental Fig. 2A), consistent with gene-level results. Using LR-Ribo-STAMP, we were able to successfully quantify EditsC for two isoforms (labeled A or B for convenience) of the same gene for 31 genes (Fig. 3A). Ordering mRNA isoforms of those 31 genes by translation levels highlighted two isoforms of *PCNP*: a protein-coding isoform with higher editing and expression than an isoform predicted to be subjected to nonsense-mediated decay (NMD) (Fig. 3B; Supplemental Fig. 2B). This aligns with NMD's association with transcript degradation

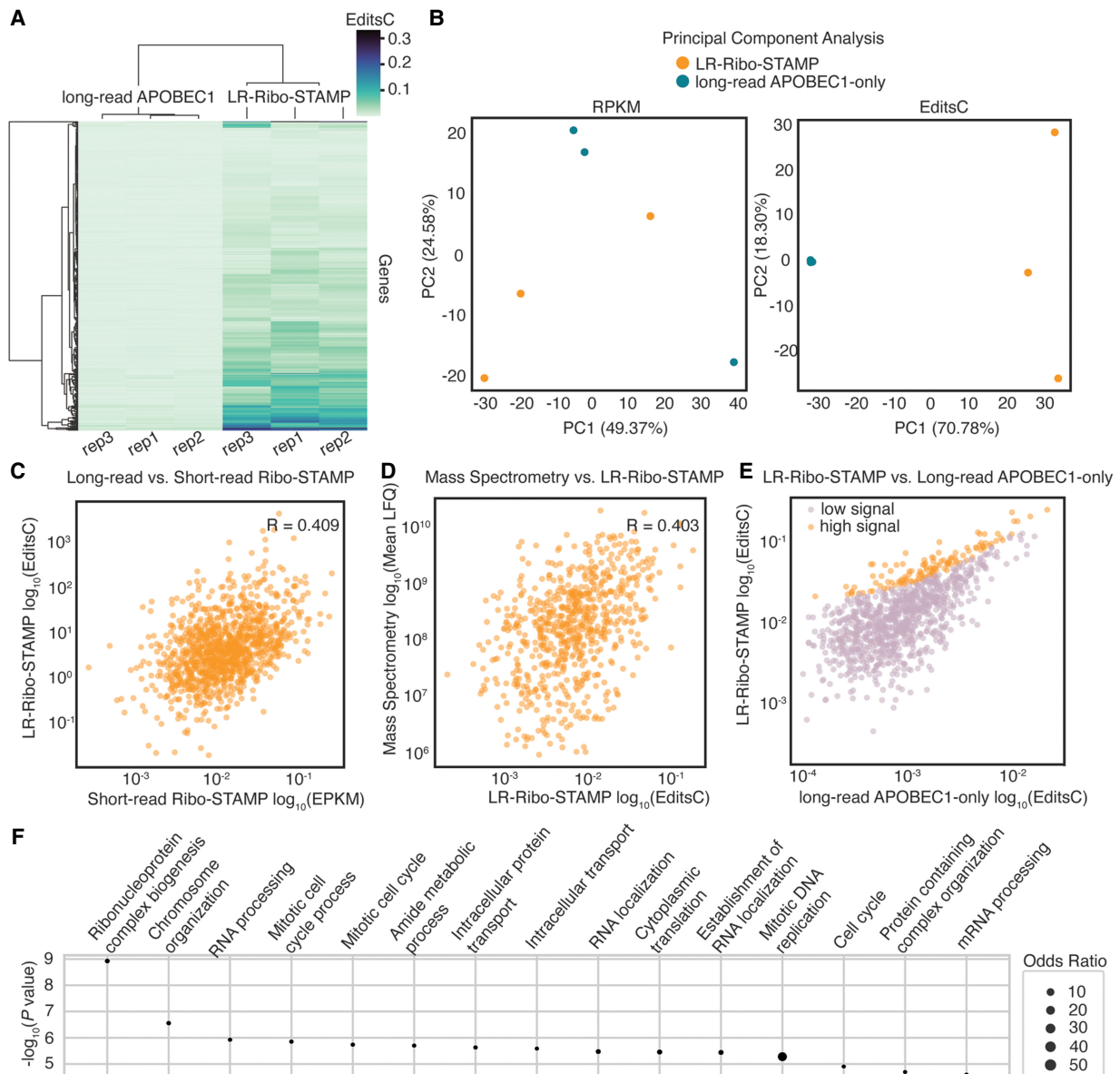


Figure 2. LR-Ribo-STAMP profiles translation at the gene level with good concordance with short-read data. (A) Signal detection (EditsC) across replicates of LR-Ribo-STAMP and long-read APOBEC1-only samples at the gene level. (B) PCA plot showing clustering of LR-Ribo-STAMP (gold) and long-read APOBEC1-only (green) samples based on gene-level RPKM (gene expression; *left*) and EditsC (editing; *right*) metrics. (C) Spearman's rank correlation of LR-Ribo-STAMP EditsC computed from long-read sequencing data and Ribo-STAMP EPKM computed from short-read sequencing data. (D) Spearman's rank correlation of LR-Ribo-STAMP and mass spectrometry data collected from unperturbed HEK293T cells. (E) Results from a linear model built using gene-level EditsC from LR-Ribo-STAMP and long-read APOBEC1-only. The results delineate genes with high signal (orange) over the background (pink). (F) Gene Ontology enrichment of genes with high signal as designated by the linear model, ranked by $-\log_{10}(P\text{-value})$.

and inhibited translation (Nickless et al. 2017) and highlights LR-Ribo-STAMP's ability to profile multiple types of mRNA isoforms.

After ranking all isoforms by LR-Ribo-STAMP EditsC, we categorized the top and bottom quartiles as high- and low-translation isoforms, respectively. Subsequently, our analysis focused on contrasting isoform features that affect translation among these groups. There were 299 genes represented in the high-translation category and 293 genes in the low-translation category. This approach was adopted as a means to confirm the efficacy of LR-Ribo-STAMP in quantifying isoform-level translation in the ab-

sence of established techniques that profile isoform-level translation with long-read sequencing. Initially, we aimed to limit the analysis to genes with multiple isoforms represented in the data set. However, this proved impractical because of the small number of isoforms and less-differentiated EditsC profiles (Supplemental Fig. 2C), which hampered meaningful statistical analysis. Therefore, we broadened our analysis to include isoforms of all genes. We placed emphasis on untranslated regions (UTRs), which are known to have regulatory significance in translation but have historically been challenging to characterize using short-read sequencing (Hinnebusch et al. 2016).

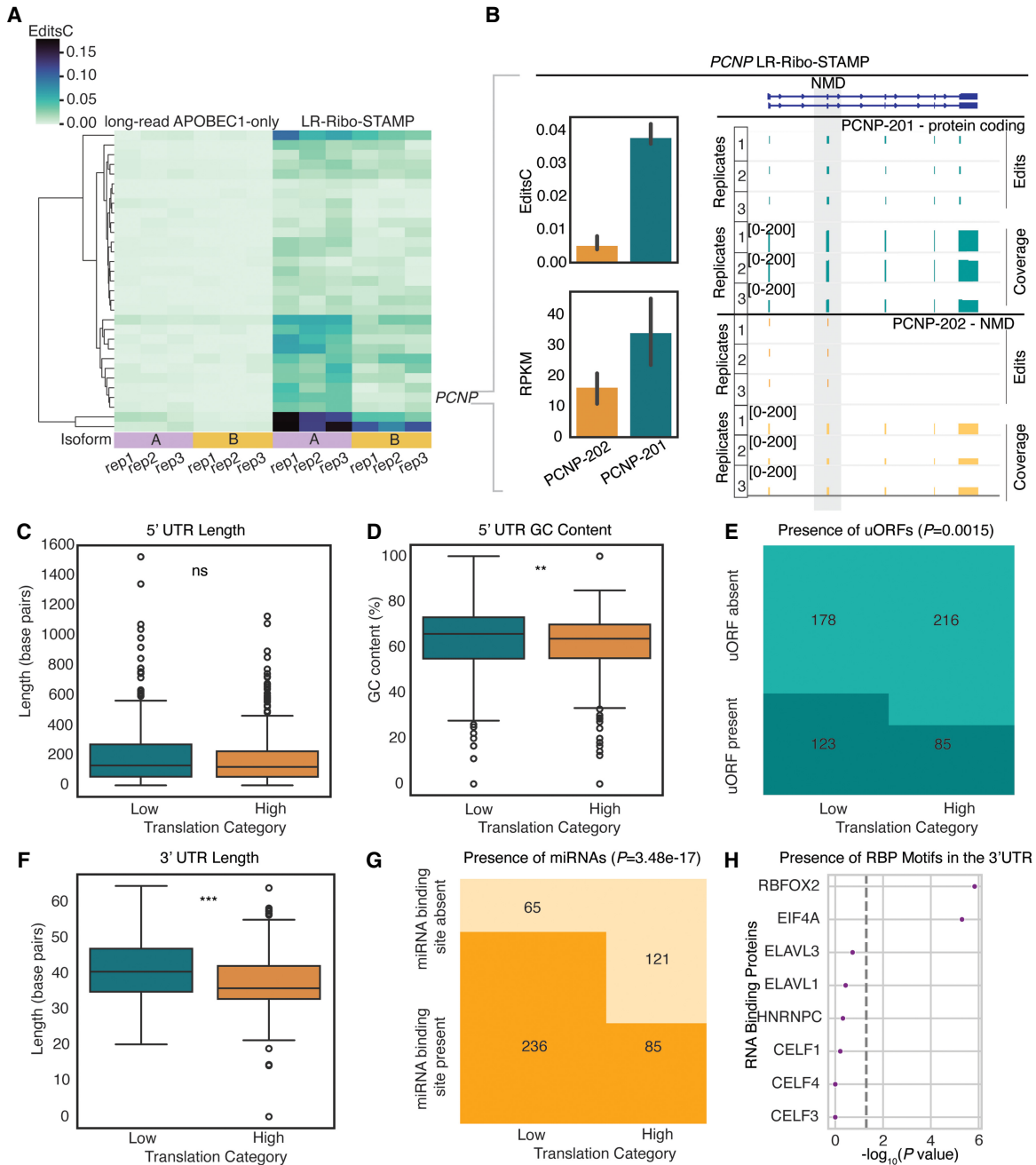


Figure 3. LR-Ribo-STAMP can profile translation at the mRNA isoform level for cells in an unperturbed state. (A) Heatmap showing EditsC quantified in long-read APOBEC1-only and LR-Ribo-STAMP samples for two mRNA isoforms of the same gene. (B) LR-Ribo-STAMP EditsC and mRNA isoform expression for two isoforms, PCNP-201 (ENST00000265260, protein-coding) and PCNP-202 (ENST00000460231, NMD), of the gene *PCNP*. (C) Comparison of 5'-UTR length of highly versus lowly translated mRNA isoforms. (D) Comparison of highly translated 5'-UTR GC content versus lowly translated mRNA isoforms. (E) The contingency table used to analyze the differences in the proportion of isoforms having uORFs between highly and lowly translated isoforms. (F) Comparison of 3'-UTR length of highly versus lowly translated isoforms. (G, H) Comparative analysis of the different proportions of isoforms having miRNA binding sites (G) and RBP motifs in the 3'-UTR sequence (H) between highly versus lowly translated isoforms. Significance: (***) $P \leq 0.001$, (**) $P \leq 0.01$, (*) $P \leq 0.05$.

In the 5' UTR, we examined differences in length, GC content, and upstream open reading frame (uORF) prevalence between high- and low-translation isoforms. Although the 5'-UTR lengths did not significantly differ (Wilcoxon rank-sum, $P = 4.11 \times 10^{-1}$), low-translation isoforms had statistically significantly higher GC content (Wilcoxon rank-sum, $P = 7 \times 10^{-3}$) (Fig. 3C,D).

Using TISdb (Wan and Qian 2014) to obtain predicted uORFs, we found that a statistically significantly higher proportion of low-translation isoforms contained predicted uORFs compared with high-translation isoforms (chi-squared test, $P = 1.5 \times 10^{-3}$) (Fig. 3E). These observations largely agree with canonical translation models (Pelletier and Sonenberg 1985; Calvo et al. 2009;

Leppke et al. 2018). In the 3' UTR, we assessed length, microRNA (miRNA) binding sites, and RNA-binding protein (RBP) binding site prevalence. Lowly translated isoforms had statistically significantly longer 3' UTRs (Wilcoxon rank-sum, $P = 7.14 \times 10^{-9}$) (Fig. 3F). This is expected as longer 3' UTRs often contain sequences recognized by regulatory elements that impact transcript stability, localization, and translation. miRNA binding sites and RBP motifs are two such features. After overlapping predicted miRNA binding sites from TargetScan (Agarwal et al. 2015) with data from LR-Ribo-STAMP, we found that a statistically significantly higher proportion (chi-squared test, $P = 3.48 \times 10^{-17}$) of lowly translated isoforms contained miRNA binding sites (Fig. 3G), in concordance with previous studies (Oliveto et al. 2017). Focusing on a subset of RBPs selected based on having potential implications in translation, EIF4A (chi-squared test, $P = 1 \times 10^{-6}$) and RBFOX2 (chi-squared test, $P = 5 \times 10^{-6}$) motifs were statistically significantly more prevalent in lowly translated isoforms (Fig. 3H; Supplemental Fig. 2D,E). In summary, LR-Ribo-STAMP is the first method to use long-read sequencing to concurrently profile mRNA translation, enhancing our ability to effectively profile translation of mRNA isoforms to extract regulatory features of translation regulation.

LR-Ribo-STAMP discovers differential gene-level translation and transcription in hypoxic conditions

We next applied LR-Ribo-STAMP to discover changes in mRNA translation upon perturbation of cellular states. We used the MDA-MB-231 triple-negative breast cancer (TNBC) cell line after 48 h of treatment with CoCl_2 , a commonly used hypoxia mimetic that blocks degradation of HIF1A, a transcription factor that regu-

lates hypoxia-inducible genes (Masoud and Li 2015; Tripathi et al. 2019). We designed the treatment to mirror the physiological conditions of prolonged hypoxia found in solid tumors in which cancers such as TNBC exhibit adaptive responses, including invasiveness and mortality (Zarrilli et al. 2020).

Our gene-level analysis identified 6242 genes having LR-Ribo-STAMP editing and at least 20 long reads in control normoxia and hypoxia conditions across replicates. We did not observe any significant global changes in translation across genes post-treatment (Wilcoxon rank-sum, $P = 1.01 \times 10^{-1}$), consistent with results from a surface sensing of translation (SUnSET) assay (Fig. 4A; Schmidt et al. 2009), with equal loading (Supplemental Fig. 3A). These observations align with previous studies showing that cancer cells are primarily glycolytic regardless of oxygen availability and, therefore, less sensitive to hypoxia than healthy cells (Shiratori et al. 2019). Correlation of gene-level LR-Ribo-STAMP EditsC across replicates shows good reproducibility and correlation of changes in expression and translation following treatment, reflecting a tight coregulation of transcription and translation (Supplemental Fig. 3B,C).

We then analyzed variations in the expression and translation of *HIF1A* between normoxia and hypoxia. The western blot showed an accumulation of HIF1A following hypoxia, in line with previous studies that have shown stabilization of HIF1A protein under hypoxia conditions (Epstein et al. 2001; Muñoz-Sánchez and Cháñez-Cárdenas 2019). Although transcriptional changes of *HIF1A* following CoCl_2 -induced hypoxia are not well studied, especially under prolonged conditions, we observed an overall decrease in expression and translation but a slight increase in TE, the ratio of EditsC to RPKM (Supplemental Fig. 3D,E). Gene

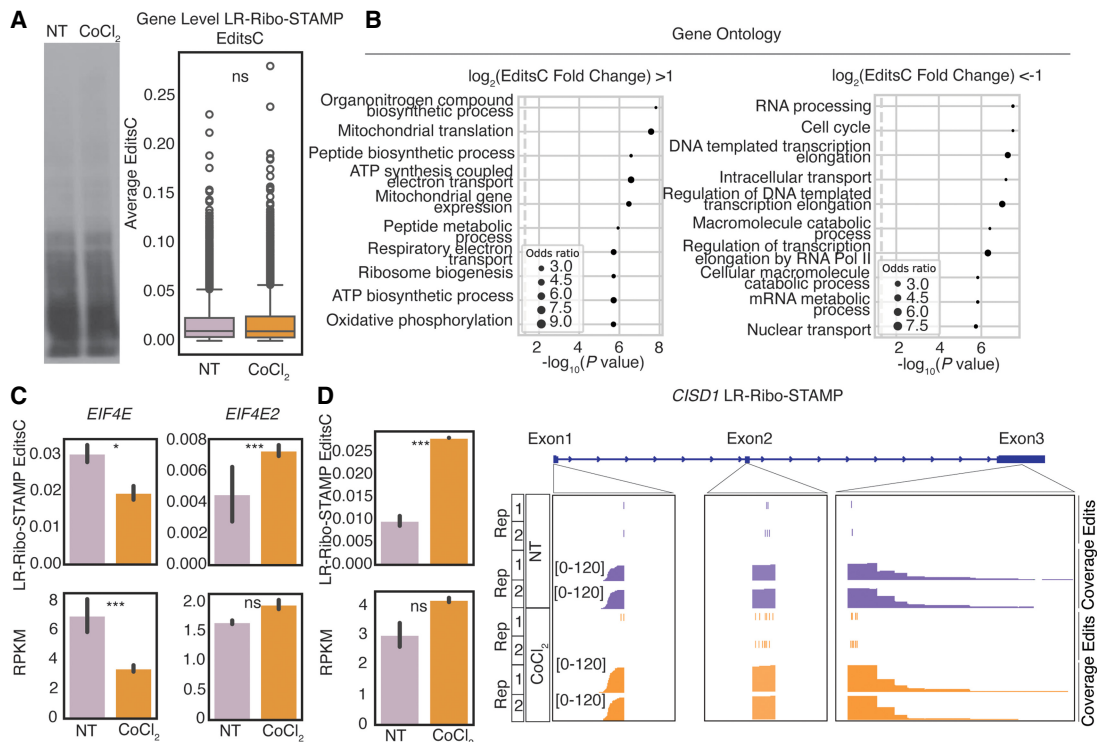


Figure 4. LR-Ribo-STAMP can profile changes in translation at the gene level for cells in disease state. (A) SUnSET assay and global quantification of LR-Ribo-STAMP EditsC at the gene level for normoxia (NT) and hypoxia (CoCl_2) treatment conditions. (B) Gene Ontology analyses of genes having higher (left) and lower (right) LR-Ribo-STAMP EditsC following hypoxia. (C) LR-Ribo-STAMP EditsC and RPKM of *EIF4E* and *EIF4E2* in normoxia and hypoxia. (D) LR-Ribo-STAMP EditsC and gene expression of *CISD1*. Significance: (***) $P \leq 0.001$, (**) $P \leq 0.01$, (*) $P \leq 0.05$.

Ontology enrichment analysis on translationally upregulated and downregulated genes showed the enrichment of categories associated with cellular adaptation to oxygen-depletion conditions (Lee et al. 2020; Adzighbli et al. 2022; Mao et al. 2024). Translationally upregulated genes reflected the use of alternate pathways like anaerobic respiration to maintain cellular energy and function. Translationally downregulated genes reflect reduced cellular activity and energy conservation (Fig. 4B).

There is a known switch in hypoxia-associated protein synthesis machinery in which hypoxia-inducible factors recruit a hypoxic complex, including EIF4E2 but not EIF4E, to the 5' cap of the 3' UTR at transcripts containing RNA hypoxia-responsive elements (Uniacke et al. 2012; Melanson et al. 2017). Therefore, we focused on *EIF4E* and *EIF4E2* to observe an example of a specific shift in cellular pathways following hypoxia. Notably, we observed the switch in expression and a statistically significant switch in translation from *EIF4E* (two-sample *t*-test, $P = 4.6 \times 10^{-2}$) to its homolog *EIF4E2* (two-sample *t*-test, $P = 3 \times 10^{-3}$) following hypoxia (Fig. 4C). We also identified that the *CISD1* gene that harbored no mRNA expression changes showed increased translation (two-sample *t*-test, $P = 3 \times 10^{-3}$) and increased TE differences (Fig. 4D; Supplemental Fig. 3F). *CISD1* has previously been shown to promote the proliferation of cancer cells, has been associated with poor survival, and has been suggested as a prognostic for breast cancer (Sohn et al. 2013; Mittler et al. 2019; Liu et al. 2023a). Overall, LR-Ribo-STAMP effectively profiles differential translation at the gene level, revealing critical shifts in regulatory molecules.

LR-Ribo-STAMP assesses changes in transcription and translation at mRNA isoform resolution in hypoxic conditions

Given that transcriptome analyses in disease models have been largely confined to the gene level, LR-Ribo-STAMP provides an avenue for discovering mRNA isoforms sensitive to changes in cellular state. We called edits across the transcriptome (Supplemental Table 2) and applied hierarchical clustering and Ward's method to identify five distinct clusters based on expression and translation changes of 5173 isoforms, all having at least 20 long reads across replicates and conditions (Fig. 5A). There were 490 genes with multiple mRNA isoforms represented in this group, although all genes were used for downstream analysis to ensure robust statistical comparisons.

Based on LR-Ribo-STAMP EditsC and differential expression analyses completed using the long-read data, we observed a strong and global positive correlation between changes in isoform expression and translation (Supplemental Fig. 4A). However, clustering enabled cluster-specific association with Gene Ontology terms and delineation of isoforms exhibiting changes in both expression and translation versus translation only. Cluster 1 showed unchanged expression but increased translation and association with respiration and electron transport chain terms, whereas cluster 2 showed increases in both and association with metabolic process terms. Cluster 3 had variable translation changes without expression alterations and association with transcription terms; cluster 4 had decreased translation with no expression change and association with cell cycle terms; and cluster 5 saw decreases in both expression and translation and association with transport and localization terms. (Fig. 5B,C). Notably, the Gene Ontology terms across clusters were consistent with cellular adaptation to oxygen-depletion conditions.

Knowing the regulatory potential of UTR regions, we examined attributes of the 5' and 3' UTRs and their associations with

translation profiles. Based on the top enriched motif for each group found by motif enrichment analysis of the 5'-UTR and 3'-UTR sequences for each cluster, we found distinct sequences associated with the different clusters of mRNA isoforms (Fig. 5D). Specifically, cluster 2's 5'-UTR motif AUUUUUUU resembled the binding site of the transcription factor, MAFF, known to be induced by HIF-1 under hypoxia conditions and to promote disease progression by increasing invasive and metastatic behavior in tumor cells (Supplemental Fig. 4B; Moon et al. 2021). The CCCAGG transcriptional motif in cluster 2, similar to those in clusters 3 and 5, resembled the motif of EBF1, a highly expressed transcription factor in TNBC cells that directly interacts with HIF1A to suppress its activity (Supplemental Fig. 4C; Qiu et al. 2022).

Additionally, we examined hypoxia-induced inhibition of the MTOR pathway (Arsham et al. 2003), which is known to translate mRNAs with 5'TOP motifs preferentially. Although we did not observe a global translation change in all isoforms with 5'TOP motifs, specific isoforms like a select one of *HSP90AB1*, regulated by mTORC1 and containing a TOP motif, showed reduced translation (Thoreen et al. 2012). We also investigated changes in the translation of isoforms containing hypoxia response elements (HREs) in the 5' UTR. Isoforms containing HREs are expected to increase in hypoxic conditions (Harris 2002). We did not observe a global change in all isoforms with HREs. However, specific isoforms, such as that of *MIF*, that have reportedly been upregulated by hypoxia in breast cancer cell lines (Bando et al. 2003) showed increased translation following hypoxia (Fig. 5E). The lack of global changes seen in isoforms containing 5'TOP motifs or HRE elements likely has to do with cancer cells already being in a glycolytic state, as mentioned before.

Lastly, we explored the role of AS, intending to connect changes in the transcriptome to changes in the transcriptome. AS is a mechanism centrally placed between transcription and translation and can determine transcriptome and transcriptome complexity through the inclusion or exclusion of exons and introns. AS analysis using the LR-Ribo-STAMP data revealed alternative transcription termination site (ATTS) enrichment between normoxic and hypoxic conditions (Fig. 5F,G). In our examination of genes that demonstrate changes in ATTS following hypoxia, we focused on *GRK6*. *GRK6* is a member of the G protein-coupled receptor kinase (GRK) family previously implicated in inducing hypoxia-inducible factor (HIF) activity in lung adenocarcinoma (Yao et al. 2021). Our analysis uncovered previously unrecognized significant shifts in translation of the isoforms resulting from varying ATTS usage following hypoxia. The ATTS usage manifests as two distinct mRNA isoforms: the shorter GRK6-206 (ENST00000507633) and longer GRK6-201 (ENST00000355472). Under hypoxic conditions, we predominantly see enhanced expression and translation of GRK6-206, whereas GRK6-201 sees a reduction in both (Fig. 5H; Supplemental Fig. 4D). We confirmed this result by western blot analysis (Supplemental Table 3) that shows increased protein abundance of the shorter GRK6-206 protein isoform in comparison to the longer GRK6-201 protein after 48 h of hypoxia (Fig. 5I). Notably, the GRK6-206 isoform lacks an AGC kinase domain when translated. The AGC kinase domain is critical for GRK proteins to properly phosphorylate G protein-coupled receptors (GPCRs), which have been linked to tumor growth and metastasis (Dorsam and Gutkind 2007; Pearce et al. 2010). This example illustrates the complex interplay between AS, transcription, and translation in a disease context. Overall, LR-Ribo-STAMP effectively elucidates the relationship between independent changes in transcription and translation at mRNA isoform resolution in

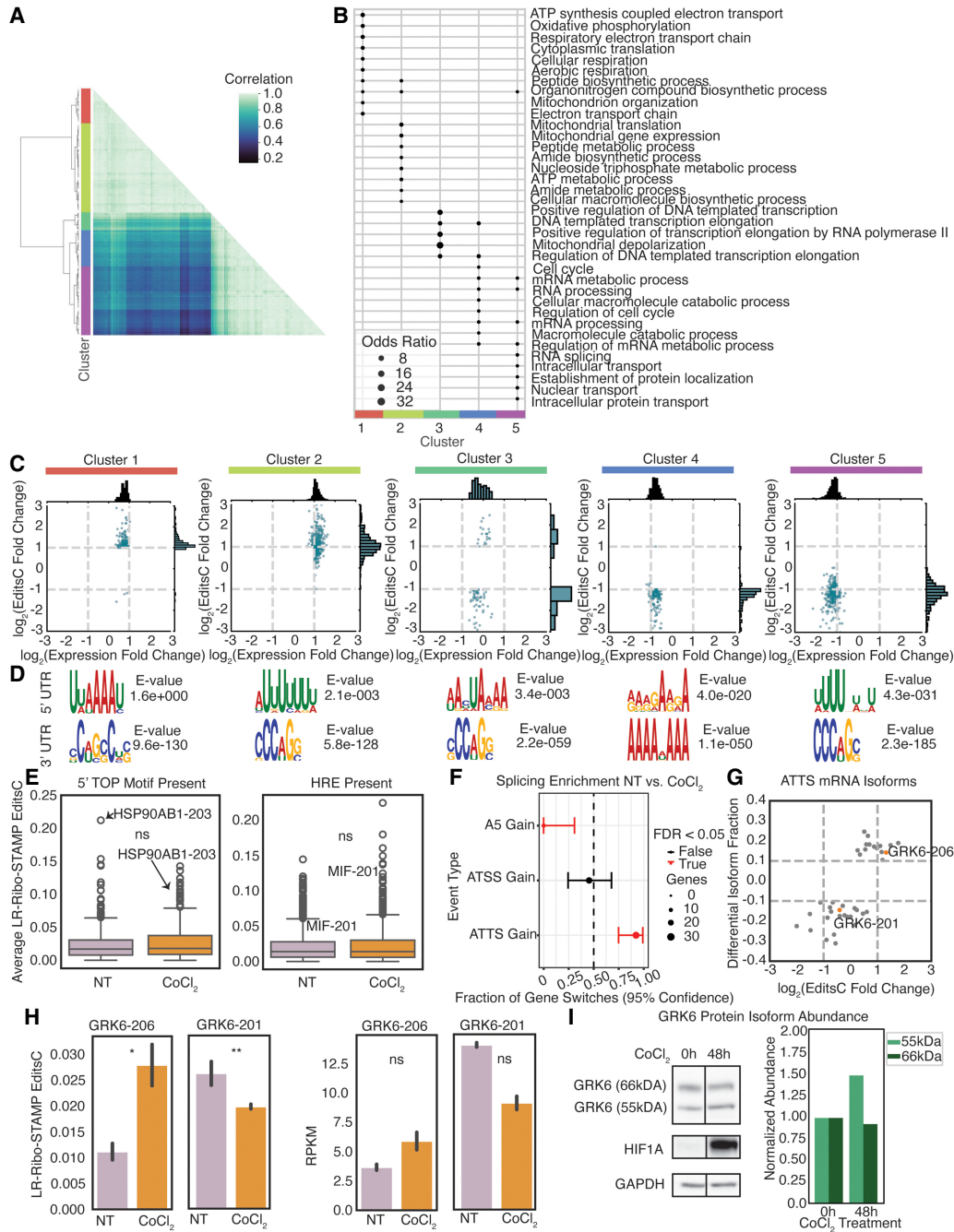


Figure 5. LR-Ribo-STAMP can profile changes in translation at the mRNA isoform level for cells in disease state. (A) Hierarchical clustering of mRNA isoforms based on LR-Ribo-STAMP EditsC and RPKM metrics. The color bar indicates correlations, and the annotations indicate clusters. (B) Gene Ontology enrichment analysis by cluster. (C) Cluster-specific changes in isoform translation ($\log_2(\text{EditsC fold change})$) versus change in expression ($\log_2(\text{expression fold change})$) following hypoxia. (D) The top enriched motif found in 5'-UTR (top) and 3'-UTR (bottom) sequences for isoforms in each cluster. (E) Translation in normoxia and hypoxia conditions for mRNA isoforms containing 5'TOP motifs (left) and HSEs (right). (F) Splicing event enrichment following hypoxia. (G) Differences in isoform fraction usage (DIF) versus change in translation following hypoxia. (H) LR-Ribo-STAMP of EditsC (left) and isoform expression (RPKM; right) of GRK6-206 (ENST00000507633) and GRK6-201 (ENST00000355472) mRNA isoforms. (I) Western blot of the protein isoforms that result from GRK6-206 (55 kDa) and GRK6-201 (66 kDa), at the 0 h and 48 h of hypoxia. Values are normalized against the 0 h timepoint. Significance: (***) $P \leq 0.001$, (**) $P \leq 0.01$, (*) $P \leq 0.05$.

disease-modeling contexts. It also points to potentially critical regulatory elements and switches in mRNA isoform transcription and translation that can inform the discovery of new mechanisms.

Discussion

Long-read sequencing platforms have enabled a level of transcriptome discovery that was previously challenging to obtain,

significantly enhancing our appreciation of the diversity of alternative mRNA isoforms (Amarasinghe et al. 2020; Marx 2023). Long-read platforms continue to improve in throughput, accuracy, and accessibility, and with the emergence of single-cell long-read sequencing, they are increasingly combined with other technologies like CRISPR-Cas9, ATAC-seq, and STAMP (Brannan et al. 2021; Hu et al. 2023; Simpson et al. 2023). This integration is unlocking new avenues to explore complex biological phenomena. Despite this, transcriptome-wide analysis of translation with full-length mRNA isoform sensitivity remains challenging owing to the incompatibility of current state-of-the-art translation profiling methods with long-read sequencing.

To address this, we developed an experimental and computational framework featuring long-read sequencing with Ribo-STAMP (LR-Ribo-STAMP) to acquire transcription and translation information with mRNA isoform resolution simultaneously. Using a specialized platform-agnostic computational pipeline to filter for signal, we showcase the effectiveness of LR-Ribo-STAMP in scalable profiling of transcription and translation at both the gene and mRNA isoform levels using RNA editing in long reads as a proxy for ribosome association (Fig. 1). We observed a positive correlation in gene-level editing quantification between Ribo-STAMP data acquired with short-read and long-read sequencing platforms, illustrating that the technology can be used to profile gene and isoform translation in unperturbed cells, and suggested that LR-Ribo-STAMP readouts may be used as a proxy for protein abundance (Figs. 2 and 3). When applied to evaluate differences in normoxia versus hypoxia states in TNBC, LR-Ribo-STAMP effectively captures variations in transcription and translation. By simultaneously profiling translation and transcription, we could link specific translation and transcriptional profiles to specific biological processes, identify critical sequence elements in UTRs, and map them to regulatory elements. By tying AS changes to mRNA isoform translation, we identified *GRK6*, which exhibited a hypoxia-induced shift to an mRNA isoform that generates a protein isoform lacking a critical protein domain, demonstrating the importance of understanding the interplay between transcription and translation (Figs. 4 and 5).

Our method represents a notable advance in the field by enabling quantification of translation at both the gene and isoform levels, a capability beyond that of established gold-standard methods and short-read Ribo-STAMP. Despite this, however, LR-Ribo-STAMP confronts challenges associated with long-read sequencing and RNA-mediated editing technology platforms. The ability to simultaneously measure translation and transcription is contingent upon having sufficient and cost-effective sequencing throughput, a hurdle yet to be fully overcome. However, recent advancements in high-throughput sequencing platforms, such as the Revio and PromethION, alongside new methods, such as PacBio's Kinnex RNA kit, which uses concatenation to increase throughput, show potential in addressing this challenge. In addition, LR-Ribo-STAMP requires accurate edit detection, which depends heavily on the accuracy of the reads and the ability to minimize background editing and biases stemming from the fused editing enzyme. Recent advancements in sequencing accuracy, such as Oxford Nanopore Technologies' Q20+ chemistry, and an expanding selection of RNA editing enzymes (Medina-Munoz et al. 2024) are helping to mitigate this issue as well. Finally, a general limitation of Ribo-STAMP stems from the need to stably integrate the construct into the genome, which currently confines its use to cell lines rather than tissues. However, with the anticipated development of an in situ meth-

od, we expect the application of this technology to extend to tissue samples in the future.

With these improvements and the continued development of more specialized computational approaches for distinguishing signal from background in Ribo-STAMP data sets and expanded isoform annotations, LR-Ribo-STAMP will be increasingly influential for profiling the translome and transcriptome complexity. This includes the ability to analyze rare and unannotated transcripts. The extensive data generated by this method are ideal for gleaning critical regulatory pathways and mechanisms and constructing context-specific translome and transcriptome profiles. Short-read Ribo-STAMP has already been coupled with short-read single-cell sequencing (Brannan et al. 2021). With advancements in long-read single-cell sequencing, there is untapped potential for LR-Ribo-STAMP to be used to profile transcriptional and translational heterogeneity at the single-cell level.

Methods

Generation of stable Ribo-STAMP and APOBEC1-only HEK293XT cell line and sequencing data

Plasmid construction, cell culture conditions and maintenance, and generation of dox-inducible HEK293XT Ribo-STAMP (RPS2-APOBEC1) and APOBEC1-only stable cell lines were completed in accordance with methods outlined by Brannan et al. (2021). For stable cell Ribo-STAMP and APOBEC1-only protein expression, cells were induced with 1 $\mu\text{g}/\text{mL}$ dox for 72 h. Total RNA was isolated from three biological replicate samples of HEK293XT cells expressing Ribo-STAMP and APOBEC1-only constructs using TRIzol extraction and column purification using the Direct-zol miniprep kit (Zymo Research). Poly(A) selection was completed using the Poly(A) mRNA magnetic isolation module (NEB E7490L), and RNA quality was assessed using high-sensitivity RNA TapeStation (Agilent 5067-5579). Long-read RNA-seq libraries were prepared using the PacBio Iso-Seq express protocol (101-763-800) and PacBio SMRTbell express template prep kit 2.0 (100-938-900). Samples were barcoded using the PacBio barcoded overhang adapter kit (101-791-700) and then pooled in an equimolar fashion. Samples were sequenced on a SMRT Cell 8M with a 30-h movie time on the PacBio Sequel II system.

Generation of stable Ribo-STAMP MDA-MB-231 cell line normoxia and induced hypoxia sequencing data

Plasmid construction of Ribo-STAMP (RPS2-APOBEC1) was completed in accordance with methods outlined by Brannan et al. (2021). MDA-MB-231s (ATCC HTB-26) were transduced with the lentiviral Ribo-STAMP vector for 24 h before treatment with puromycin (2 mg/mL). Following 48 h of puromycin selection, cells were sorted for the top 10% of mRuby-Ribo-STAMP expressing cells on a BD influx cell sorter. Cells with dox-inducible Ribo-STAMP were expanded and then cultured in DMEM + 10% FBS (Gibco) containing 1 $\mu\text{g}/\text{mL}$ dox to induce Ribo-STAMP expression and 100 μM cobalt (II) chloride (Sigma-Aldrich 15862-1ML-F) to simulate hypoxia. Following 48 h of DOX and cobalt (II) chloride treatment, cells were harvested from normoxia and induced hypoxia conditions. RNA was isolated for biological duplicate samples with TRIzol extraction and column purification using the Direct-zol miniprep kit (Zymo Research). RNA quality was assessed using RNA screen tape (Agilent 5067-5576). Poly(A) site selection was completed using the Poly(A) mRNA magnetic isolation module (NEB E7490L). Long-read RNA-seq libraries were then prepared from extracted RNA using the SMRTbell prep kit v3.0 (102-141-

700). Libraries were barcoded, pooled in an equimolar fashion, and sequenced using two SMRT Cells 8M with a 30-h movie time on the PacBio Sequel IIe.

SUnSET assay

MDA-MB-231 cells were treated with 100 μ M cobalt (II) chloride to induce hypoxia for 48 h. Cells were then treated with 10 μ g/mL puromycin for 10 min and then subsequently processed for western blot analysis. To process the samples for western blot, cells were lysed in RIPA buffer (Sigma-Aldrich) with 200 \times protease inhibitor and quantified with the Pierce BCA protein quantification kit (Thermo Fisher Scientific 23225). Lysates were run on a 4%–12% NuPAGE Bis-Tris gel in NuPAGE MOPS running buffer (Thermo Fisher Scientific) and transferred to a polyvinylidene fluoride (PVDF) membrane. The membrane was first incubated in Ponceau S stain to obtain total protein staining for SUnSET assay normalization. Then, the membrane was blocked in 5% nonfat milk in TBST for 30 min and incubated overnight at 4°C with the mouse antipuro-mycin (clone 12D10, Millipore Sigma MABE343) antibody. The membrane was washed three times for 5 min each time in TBST, incubated for 1 h at room temperature in 5% nonfat milk in TBST with a horse radish peroxidase-conjugated antimouse secondary antibody (Cell Signaling Technology 7076), and washed three times for 5 min each time in TBST and developed using Pierce ECL western blotting substrate (Thermo Fisher Scientific 32132).

Western blot

MDA-MB-231 cells were treated with 100 μ M cobalt chloride II to simulate hypoxia for 24 or 48 h and then lysed with RIPA buffer (Sigma-Aldrich) containing Protease inhibitor (Thermo Fisher Scientific). Protein lysates were centrifuged to pellet and remove insoluble material and were then quantified using the Pierce BCA kit. Protein lysates were run on a 4%–12% NuPAGE Bis-Tris gel and transferred to a PVDF membrane. Membranes were blocked in Tris-buffered saline containing Tween 20 (TBST) with 5% milk for 20 min and probed overnight at 4°C with primary antibody (rabbit pAB anti GRK6 [N terminal], Abcam ab244364; rabbit mAB anti HIF-1 α , Cell Signaling Technology 14179; mouse mAB anti GAPDH, Millipore MAB374). Membranes were washed three times for 5 min with TBST and then probed for 1 h at room temperature in TBST containing 5% milk with secondary antibody (antimouse IgG, HRP linked, Cell Signaling Technology 7076; antirabbit IgG, HRP linked, Cell Signaling Technology 7074) diluted 1:5000. Membranes were washed three times for 5 min with TBST and developed using Thermo Pierce ECL detection kits on an Azure western blot imaging system.

RNA-seq data processing, QC, and generation of count matrices and isoform read assignments

All data processing was completed using the Triton Shared Computing Cluster (<https://doi.org/10.57873/T34W2R>). Demultiplexed circular consensus sequence (CCS) reads obtained after sequencing were processed using the Iso-Seq v4 pipeline (Epstein et al. 2001; <https://isoseq.how/>). First, full-length nonconcatemer reads were generated using lima v2.9.0 with the parameter `--isoseq`. Reads were then refined with Iso-Seq v4.0.0's refine tool with the parameter `--require-polya`. Refined reads from HEK293XT and MDA-MB-231 samples were aligned to the GRCh37 and GRCh38 reference genomes, respectively, using pbmm2 v1.13.1 align with the parameter `--preset ISOSEQ`. HEK293XT samples were aligned to GRCh37 to maintain consistency and comparability with analyses completed by Brannan et al. (2021). GRCh38 was

used to leverage the most updated, comprehensive, and widely accepted reference for this proof-of-concept study.

The quality of aligned reads was assessed using NanoPlot v1.32.1 (De Coster and Rademakers 2023) with the parameters `--raw` and `--tsv_stats`. Reads with a quality score of less than 20, as assessed by NanoPlot, along with unmapped reads, supplemental alignment reads, secondary alignment reads, and those aligned to the incorrect strand, were excluded from the analysis (`filter_bam_v2.py`). Following read filtering, mRNA isoform-level count matrices and read assignments were obtained using IsoQuant v3.3.0 (Prijbelski et al. 2023) with the parameters `--data_type pacbio`, `--transcript_quantification unique_only`, and `--gene_quantification unique_only`. Reference genome GRCh37 and GENCODE comprehensive annotation GRCh37 (v19) were used for generating isoform counts for HEK293XT sample data, and reference genome GRCh38 and GENCODE comprehensive annotation GRCh38 (v38) were used for generating isoform counts for MDA-MB-231 samples. Read assignment (`read_assignment.tsv`) output from IsoQuant was used to assign individual mapped reads to isoforms. Count matrices were used to calculate RPKM mapped reads. The RPKM was calculated in accordance with the following equation: $RPKM = \text{number of reads} / (\text{gene length} / 1000 \times \text{total reads} / 1000000)$. Mapped reads were determined using SAMtools v1.16 view (Daneczek et al. 2021) with the parameter `--count`. Only genes and mRNA isoforms having at least 20 reads across each replicate in each condition were considered for downstream analysis.

Edit detection from aligned sequencing data

To facilitate isoform-specific edit detection and allow for multipro-cessing, each sample's aligned reads were divided into smaller groups. Each group contained a unique set of reads corresponding to one isoform of all genes, as assigned by the output of IsoQuant (`split_bam_isoquant.py`). Subsequently, the pileup method in pysam v0.21.0 was used to iterate through every base of every isoform to determine the count of reads at a position containing a C-to-U edit and the total number of reads at those positions (`read_level_quant_se_ct_annotated.py`). Edits are associated with one of four categories: the full transcript, 5' UTR, 3' UTR, or the CDS. The coordinates of these regions were determined using the GENCODE comprehensive annotations for GRCh37 (v19) and GRCh38 (v38) for the HEK293XT and MDA-MB-231 samples, respectively. Using the pileup method, we also associate an edited position with a read identifier.

In this study, we focused exclusively on edits within the CDS. We identified and removed edits that were present in all replicates and conditions of each sample group. Additionally, edits overlapping with positions listed in the dbSNP database (Sherry et al. 2001), corresponding to the reference genome used, were also excluded from the analysis. The remaining edited positions were considered for downstream analysis (`filter_edits_calc_editsC.py`).

Edit fraction, EditsC, and TE metrics for quantification

To calculate the edit fraction at each position, the formula used is $\text{edit fraction} = \text{edited reads} / \text{total reads}$. EditsC represents the proportion of cytosines in a gene or mRNA isoform that undergoes C-to-U editing. For determining the total count of cytosines, exon and UTR coordinates were curated at the gene and mRNA isoform levels using BEDTools (Quinlan and Hall 2010) v2.29.2 merge with the parameters `-s` and `-c 4,6`. The sequences corresponding to these regions were then obtained using BEDTools v2.29.2 `getfasta` with the appropriate reference FASTA files and the parameters `-name` and `-s`. The total number of cytosines was obtained for each gene or isoform by counting the cytosines in

the sequences. To calculate the TE of a gene or isoform, the formula used is $TE = EditsC / RPKM$.

Downstream analysis of edit quantification

Gene Ontology enrichment analyses were conducted using decoupleR v1.5.0 (Badia-i-Mompel et al. 2022) and the biological processes category. Replicate correlations were completed using Pearson's R correlation via SciPy v1.11.4 (<https://docs.scipy.org/doc/scipy-1.11.4/index.html>), whereas different sample type correlations were completed using Spearman's rank correlation with the same package. Visualization of editing, gene expression, mRNA isoform expression, and reference annotations was done using the Integrative Genomics Viewer (IGV) v2.14.1 (Robinson et al. 2011).

For HEK293T cell samples, linear regression at the gene and mRNA isoform levels was performed using statsmodels v0.14.0 (<https://www.statsmodels.org/stable/release/version0.14.0.html>), with a standard deviation threshold of 1 for residual values. For comparative analysis of short-read and long-read Ribo-STAMP data, short-read Ribo-STAMP data were acquired from the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE155729. To compare LR-Ribo-STAMP EditsC values with mass spectrometry data, we obtained label-free quantification (LFQ) intensity values from the LR-Ribo-STAMP ProteomeXchange Consortium (<https://www.proteomexchange.org>) under data set identifier PXD020630. We calculated the average LFQ intensity values across all wild-type HEK293T samples to compare against EditsC values computed from LR-Ribo-STAMP. Following this, high- and low-translation isoforms were categorized based on LR-Ribo-STAMP EditsC, with the top and bottom quartiles corresponding to high and low translation. UTR lengths were derived from GENCODE comprehensive annotation GRCh37 (v19), and Wilcoxon rank-sum (SciPy v1.11.4) was used to assess group differences. Overlaps of 5'-UTR sequences with predicted uORFs obtained from TISdb (Wan and Qian 2014) and 3'-UTR sequences with predicted binding sites from TargetScan (Agarwal et al. 2015) were identified using BEDTools v2.29.2 intersect with default parameters. Published RBP motifs (Riley et al. 2014) were obtained, and exact matches were searched for in 3'-UTR sequences. Chi-squared tests were implemented using SciPy v1.11.4.

For MDA-MB-231 samples, a two-tailed *t*-test (SciPy 1.11.4) was used to compare normoxia and hypoxia conditions at gene and isoform levels. Significant changes in translation were designated as genes or isoforms with $P \leq 0.05$ and $|\log_2(\text{EditsC hypoxia} / \text{EditsC normoxia})| \geq 1$. Following differential gene and isoform expression analysis with DESeq2 v1.39.3 with count matrices obtained from IsoQuant, significant changes in expression were designated as genes or isoforms having an adjusted $P \leq 0.05$ and $|\log_2(\text{expression hypoxia} / \text{expression normoxia})| \geq 1$. Clustering of mRNA isoforms based on transcription and translation measurements was completed using Ward's method. UTR sequences of each cluster were analyzed for enriched motif sequences using MEME v5.3.0 (Bailey and Elkan 1994). TomTom v5.5.5 (Gupta et al. 2007) was used to identify known motifs with strong similarity to those identified in the clusters. The presence or absence of sequence elements such as 5'TOP motif and HRE was determined by looking for exact sequence matches. Differences in LR-Ribo-STAMP EditsC between normoxia and hypoxia for each group were determined based on Wilcoxon rank-sum, implemented with Python package SciPy 1.11.4. Alternative splicing analysis was performed on normoxia and hypoxia sequencing data using R package (R Core Team 2023) IsoformSwitchAnalyzer v2.2.0 (Vitting-Seerup and Sandelin 2019).

Software availability

Source code and analysis scripts for edit quantification are available at GitHub (<https://github.com/YeoLab/LR-Ribo-STAMP>) and as Supplemental Code.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE255844.

Competing interest statement

G.W.Y. is a cofounder, a member of the Board of Directors, a member of the SAB, an equity holder, and a paid consultant for Locanabio (through December 31, 2023) and Eclipse BioInnovations. G.W.Y. is a distinguished 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. J.G.U. is an employee and shareholder for PacBio. J.G.U. is on the SAB and is a shareholder for Eclipse BioInnovations. The remaining authors have no competing interests to declare.

Acknowledgments

We thank members of the Yeo laboratory for helpful discussions and suggestions during the preparation of this manuscript. This work was supported by U.S. National Institutes of Health (NIH) grants HG004659, HG011864, HG009889, and MH126719 to G.W.Y. P.J. was supported by NIH institutional training grant T32 GM145427 and an ARCS scholarship. K.W.B. is supported by NIH/National Institute of Neurological Disorders and Stroke (NINDS) K22 NS112678, NIH/National Cancer Institute (NCI) R01 CA284315, and Cancer Prevention and Research Institute of Texas (CPRIT) award RR220017. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing a Sequel II.

Author contributions: Idea conception was by P.J., D.A.L., K.W.B., and G.W.Y. The methodology was developed by P.J., D.A.L., B.A.Y., and G.W.Y. Data generation of LR-Ribo-STAMP data in the HEK293T and MDA-MB-231 cell lines was completed by D.A.L., J.G.U., A.T.T., and T.Y. SUNSET assay and western blot of HIF1A were completed by A.T.T. Western blot of GRK6 was completed by C.J.Z. All bioinformatic analysis was completed by P.J. The original manuscript was written by P.J. and G.W.Y. Funding acquisition was completed by G.W.Y. All authors have read and accepted the final version of the manuscript.

References

- Adzigbli L, Sokolov EP, Wimmers K, Sokolova IM, Ponsuksili S. 2022. Effects of hypoxia and reoxygenation on mitochondrial functions and transcriptional profiles of isolated brain and muscle porcine cells. *Sci Rep* **12**: 19881. doi:10.1038/s41598-022-24386-0
- Agarwal V, Bell GW, Nam J-W, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**: e05005. doi:10.7554/eLife.05005
- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**: 30. doi:10.1186/s13059-020-1935-5
- Arsham AM, Howell JJ, Simon MC. 2003. A novel hypoxia-inducible factor-independent hypoxic response regulating mammalian target of rapamycin and its targets. *J Biol Chem* **278**: 29655–29660. doi:10.1074/jbc.M212770200

- Badia-i-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, Taus P, Dugourd A, Holland CH, Ramirez Flores RO, et al. 2022. Decoupler: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv* **2**: vbac016. doi:10.1093/bioadv/vbac016
- Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28–36.
- Bando H, Toi M, Kitada K, Koike M. 2003. Genes commonly upregulated by hypoxia in human breast cancer cells MCF-7 and MDA-MB-231. *Biomed Pharmacother* **57**: 333–340. doi:10.1016/s0753-3322(03)00098-2
- Brannan KW, Chaim IA, Marina RJ, Yee BA, Kofman ER, Lorenz DA, Jagannatha P, Dong KD, Madrigal AA, Underwood JG, et al. 2021. Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. *Nat Methods* **18**: 507–519. doi:10.1038/s41592-021-01128-0
- Calvo SE, Pagliarini DJ, Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci* **106**: 7507–7512. doi:10.1073/pnas.0810916106
- Cui H, Hu H, Zeng J, Chen T. 2019. DeepShape: estimating isoform-level ribosome abundance and distribution with Ribo-seq data. *BMC Bioinformatics* **20**: 678. doi:10.1186/s12859-019-3244-0
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**: giab008. doi:10.1093/gigascience/giab008
- De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**: btad311. doi:10.1093/bioinformatics/btad311
- Dorsam RT, Gutkind JS. 2007. G-protein-coupled receptors and cancer. *Nat Rev Cancer* **7**: 79–94. doi:10.1038/nrc2069
- Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, et al. 2001. *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**: 43–54. doi:10.1016/s0092-8674(01)00507-4
- Flamand MN, Ke K, Tamming R, Meyer KD. 2022. Single-molecule identification of the target RNAs of different RNA binding proteins simultaneously in cells. *Genes Dev* **36**: 1002–1015. doi:10.1101/gad.349983.122
- Floor SN, Doudna JA. 2016. Tunable protein synthesis by transcript isoforms in human cells. *Life* **5**: e10921. doi:10.7554/eLife.10921
- Foord C, Hsu J, Jarroux J, Hu W, Belchikov N, Pollard S, He Y, Joglekar A, Tilgner HU. 2023. The variables on RNA molecules: concert or cacophony? answers in long-read sequencing. *Nat Methods* **20**: 20–24. doi:10.1038/s41592-022-01715-9
- Gunawardana Y, Niranjan M. 2013. Bridging the gap between transcriptome and proteome measurements identifies post-translationally regulated genes. *Bioinformatics* **29**: 3060–3066. doi:10.1093/bioinformatics/btt537
- Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biol* **8**: R24. doi:10.1186/gb-2007-8-2-r24
- Harris AL. 2002. Hypoxia: a key regulatory factor in tumour growth. *Nat Rev Cancer* **2**: 38–47. doi:10.1038/nrc704
- Hegazi S, Cheng AH, Krupp JJ, Tasaki T, Liu J, Szulc DA, Ling HH, Rios Garcia J, Seecharran S, Basiri T, et al. 2022. UBR4/POE facilitates secretory trafficking to maintain circadian clock synchrony. *Nat Commun* **13**: 1594. doi:10.1038/s41467-022-29244-1
- Hinnebusch AG, Ivanov IP, Sonenberg N. 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**: 1413–1416. doi:10.1126/science.aad9868
- Hu Y, Jiang Z, Chen K, Zhou Z, Zhou X, Wang Y, Yang J, Zhang B, Wen L, Tang F. 2023. scNanoATAC-seq: a long-read single-cell ATAC sequencing method to detect chromatin accessibility and genetic variants simultaneously within an individual cell. *Cell Res* **33**: 83–86. doi:10.1038/s41422-022-00730-x
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223. doi:10.1126/science.1168978
- Kornbliht AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**: 153–165. doi:10.1038/nrm3525
- Lee P, Chandel NS, Simon MC. 2020. Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nat Rev Mol Cell Biol* **21**: 268–283. doi:10.1038/s41580-020-0227-y
- Leppek K, Das R, Barna M. 2018. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **19**: 158–174. doi:10.1038/nrm.2017.103
- Li JJ, Chew G-L, Biggin MD. 2019. Quantitative principles of cis-translational control by general mRNA sequence features in eukaryotes. *Genome Biol* **20**: 162. doi:10.1186/s13059-019-1761-9
- Lin Y, Kwok S, Hein AE, Thai BQ, Alabi Y, Ostrowski MS, Wu K, Floor SN. 2023. RNA molecular recording with an engineered RNA deaminase. *Nat Methods* **20**: 1887–1899. doi:10.1038/s41592-023-02046-z
- Liu F, Dong Y, Zhong F, Guo H, Dong P. 2023a. CISD1 is a breast cancer prognostic biomarker associated with diabetes mellitus. *Biomolecules* **13**: 37. doi:10.3390/biom13010037
- Liu Z, Quinones-Valdez G, Fu T, Huang E, Choudhury M, Reese F, Mortazavi A, Xiao X. 2023b. L-GIREMI uncovers RNA editing sites in long-read RNA-seq. *Genome Biol* **24**: 171. doi:10.1186/s13059-023-03012-w
- Mao Y, Zhang J, Zhou Q, He X, Zheng Z, Wei Y, Zhou K, Lin Y, Yu H, Zhang H, et al. 2024. Hypoxia induces mitochondrial protein lactylation to limit oxidative phosphorylation. *Cell Res* **34**: 13–30. doi:10.1038/s41422-023-00864-6
- Marx V. 2023. Method of the year: long-read sequencing. *Nat Methods* **20**: 6–11. doi:10.1038/s41592-022-01730-w
- Masoud GN, Li W. 2015. HIF-1 α pathway: role, regulation and intervention for cancer therapy. *Acta Pharm Sin B* **5**: 378–389. doi:10.1016/j.apbsb.2015.05.007
- Medina-Munoz HC, Kofman E, Jagannatha P, Boyle EA, Yu T, Jones KL, Mueller JR, Lykins GD, Doudna AT, Park SS, et al. 2024. Expanded palette of RNA base editors for comprehensive RBP-RNA interactome studies. *Nat Commun* **15**: 875. doi:10.1038/s41467-024-45009-4
- Melanson G, Timpano S, Uniacke J. 2017. The eIF4E-directed hypoxic cap-dependent translation machinery reveals novel therapeutic potential for cancer treatment. *Oxid Med Cell Longev* **2017**: e6098107. doi:10.1155/2017/6098107
- Mitschka S, Mayr C. 2022. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat Rev Mol Cell Biol* **23**: 779–796. doi:10.1038/s41580-022-00507-5
- Mittler R, Darash-Yahana M, Sohn YS, Bai F, Song L, Cabantchik IZ, Jennings PA, Onuchic JN, Nechushtai R. 2019. NEET proteins: a new link between iron metabolism, reactive oxygen species, and cancer. *Antioxid Redox Signal* **30**: 1083–1095. doi:10.1089/ars.2018.7502
- Moon EJ, Mello SS, Li CG, Chi J-T, Thakkar K, Kirkland JG, Lagory EL, Lee JJ, Diep AN, Miao Y, et al. 2021. The HIF target MAFF promotes tumor invasion and metastasis through IL11 and STAT3 signaling. *Nat Commun* **12**: 43081. doi:10.1038/s41467-021-24631-6
- Muñoz-Sánchez J, Cháñez-Cárdenas ME. 2019. The use of cobalt chloride as a chemical hypoxia model. *J Appl Toxicol* **39**: 556–570. doi:10.1002/jat.3749
- Nickless A, Bailis JM, You Z. 2017. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci* **7**: 26. doi:10.1186/s13578-017-0153-7
- Oliveto S, Mancino M, Manfrini N, Biffo S. 2017. Role of microRNAs in translation regulation and cancer. *World J Biol Chem* **8**: 45–56. doi:10.4331/wjbc.v8.i1.45
- Pearce LR, Komander D, Alessi DR. 2010. The nuts and bolts of AGC protein kinases. *Nat Rev Mol Cell Biol* **11**: 9–22. doi:10.1038/nrm2822
- Pelletier J, Sonenberg N. 1985. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* **40**: 515–526. doi:10.1016/0092-8674(85)90200-4
- Prijbelski AD, Mikheenko A, Joglekar A, Smetanin A, Jarroux J, Lapidus AL, Tilgner HU. 2023. Accurate isoform discovery with IsoQuant using long reads. *Nat Biotechnol* **41**: 915–918. doi:10.1038/s41587-022-01565-y
- Qiu Z, Guo W, Dong B, Wang Y, Deng P, Wang C, Liu J, Zhang Q, Grosschedl R, Yu Z, et al. 2022. EBF1 promotes triple-negative breast cancer progression by surveillance of the HIF1 α pathway. *Proc Natl Acad Sci* **119**: e2119518119. doi:10.1073/pnas.2119518119
- Quattrone A, Dassi E. 2019. The architecture of the human RNA-binding protein regulatory network. *iScience* **21**: 706–719. doi:10.1016/j.isci.2019.10.058
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- R Core Team. 2023. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Reixachs-Solé M, Ruiz-Orera J, Albà MM, Eyrales E. 2020. Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nat Commun* **11**: 1768. doi:10.1038/s41467-020-15634-w
- Riley TR, Slattery M, Abe N, Rastogi C, Liu D, Mann RS, Bussemaker HJ. 2014. SELEX-seq: a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods Mol Biol* **1196**: 255–278. doi:10.1007/978-1-4939-1242-1_16

- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26. doi:10.1038/nbt.1754
- Schmidt EK, Clavarino G, Ceppi M, Pierre P. 2009. SUnSET, a nonradioactive method to monitor protein synthesis. *Nat Methods* **6**: 275–277. doi:10.1038/nmeth.1314
- Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**: 308–311. doi:10.1093/nar/29.1.308
- Shiratori R, Furuichi K, Yamaguchi M, Miyazaki N, Aoki H, Chibana H, Ito K, Aoki S. 2019. Glycolytic suppression dramatically changes the intracellular metabolic profile of multiple cancer cell lines in a mitochondrial metabolism-dependent manner. *Sci Rep* **9**: 18699. doi:10.1038/s41598-019-55296-3
- Simpson BP, Yrigollen CM, Izda A, Davidson BL. 2023. Targeted long-read sequencing captures CRISPR editing and AAV integration outcomes in brain. *Mol Ther* **31**: 760–773. doi:10.1016/j.ymthe.2023.01.004
- Sohn Y-S, Tamir S, Song L, Michaeli D, Matouk I, Conlan AR, Harir Y, Holt SH, Shulaev V, Paddock ML, et al. 2013. NAF-1 and mitoNEET are central to human breast cancer proliferation by maintaining mitochondrial homeostasis and promoting tumor growth. *Proc Natl Acad Sci* **110**: 14676–14681. doi:10.1073/pnas.1313198110
- Sterne-Weiler T, Martinez-Nunez RT, Howard JM, Cvitovik I, Katzman S, Tariq MA, Pourmand N, Sanford JR. 2013. Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res* **23**: 1615–1623. doi:10.1101/gr.148585.112
- Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**: 109–113. doi:10.1038/nature11083
- Tripathi VK, Subramaniyan SA, Hwang I. 2019. Molecular and cellular response of co-cultured cells toward cobalt chloride (CoCl₂)-induced hypoxia. *ACS Omega* **4**: 20882–20893. doi:10.1021/acsomega.9b01474
- Uniacke J, Holterman CE, Lachance G, Franovic A, Jacob MD, Fabian MR, Payette J, Holcik M, Pause A, Lee S. 2012. An oxygen-regulated switch in the protein synthesis machinery. *Nature* **486**: 126–129. doi:10.1038/nature11055
- Vitting-Seerup K, Sandelin A. 2019. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**: 4469–4471. doi:10.1093/bioinformatics/btz247
- Wan J, Qian S-B. 2014. TISdb: a database for alternative translation initiation in mammalian cells. *Nucleic Acids Res* **42**: D845–D850. doi:10.1093/nar/gkt1085
- Yao S, Ertay A, Zhou Y, Yao L, Hill C, Chen J, Guan Y, Sun H, Ewing RM, Liu Y, et al. 2021. GRK6 depletion induces HIF activity in lung adenocarcinoma. *Front Oncol* **11**: 654812. doi:10.3389/fonc.2021.654812
- Zarrilli G, Businello G, Dieci MV, Paccagnella S, Carraro V, Cappellesso R, Miglietta F, Griguolo G, Guarneri V, Lo Mele M, et al. 2020. The tumor microenvironment of primitive and metastatic breast cancer: implications for novel therapeutic strategies. *Int J Mol Sci* **21**: 8102. doi:10.3390/ijms21218102

Received February 22, 2024; accepted in revised form May 31, 2024.