# 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], Gene W. Yeo[1,2,3,4,5,#]

[1] Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA
[2] Sanford Stem Cell Institution Innovation Center and Stem Cell Program, University of California San Diego, La Jolla, CA, USA
[3] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA
[4] Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, USA
[5] Sanford Laboratories for Innovative Medicine, La Jolla, CA, USA
[6] PacBio, Menlo Park, CA, USA
[7]Present address: Center for RNA Therapeutics, Department of Cardiovascular Sciences, Houston Methodist Research Institute, Houston, USA
[#]Correspondence: geneyeo@ucsd.edu

## ABSTRACT

Transcription and translation are intertwined processes where 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 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 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*, where 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 measure mRNA translation with isoform sensitivity.


## INTRODUCTION

Post-transcriptional processes such as alternative splicing 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 & 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 due to fragmentation and high input material requirements, respectively. Both approaches also require the separate generation of mRNA-seq libraries in parallel with ribosome-protected fragments (in Ribo-seq) or polysome-fractionated mRNAs (in polysome profiling) to compute translation efficiency (TE). While computational attempts have been proposed to better understand the components that affect translation (Cui et al., 2019; Gunawardana & Niranjan, 2013; Li et al., 2019; Quattrone & 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 & 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 due 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 single-cell capture approaches to successfully measure translatomes 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 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

### Long-read 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 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 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. While 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, where 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 fraction—the ratio of edited to total reads at a position (Z. Liu et al., 2023). 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 where data is sparse. Our approach leverages a read threshold that only considers positions with greater than 20 reads (**Supplemental Figure 1A**) to ensure reliable edit detection and eliminates edited sites found across sample types and replicates, reasoning that they would be unlikely due to the transient nature of ribosome association with mRNA transcripts. Testing this hypothesis with 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 Figure 1B**). While our approach does not explicitly filter based on the fraction of reads edited at a site, 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 Figure 1C**).

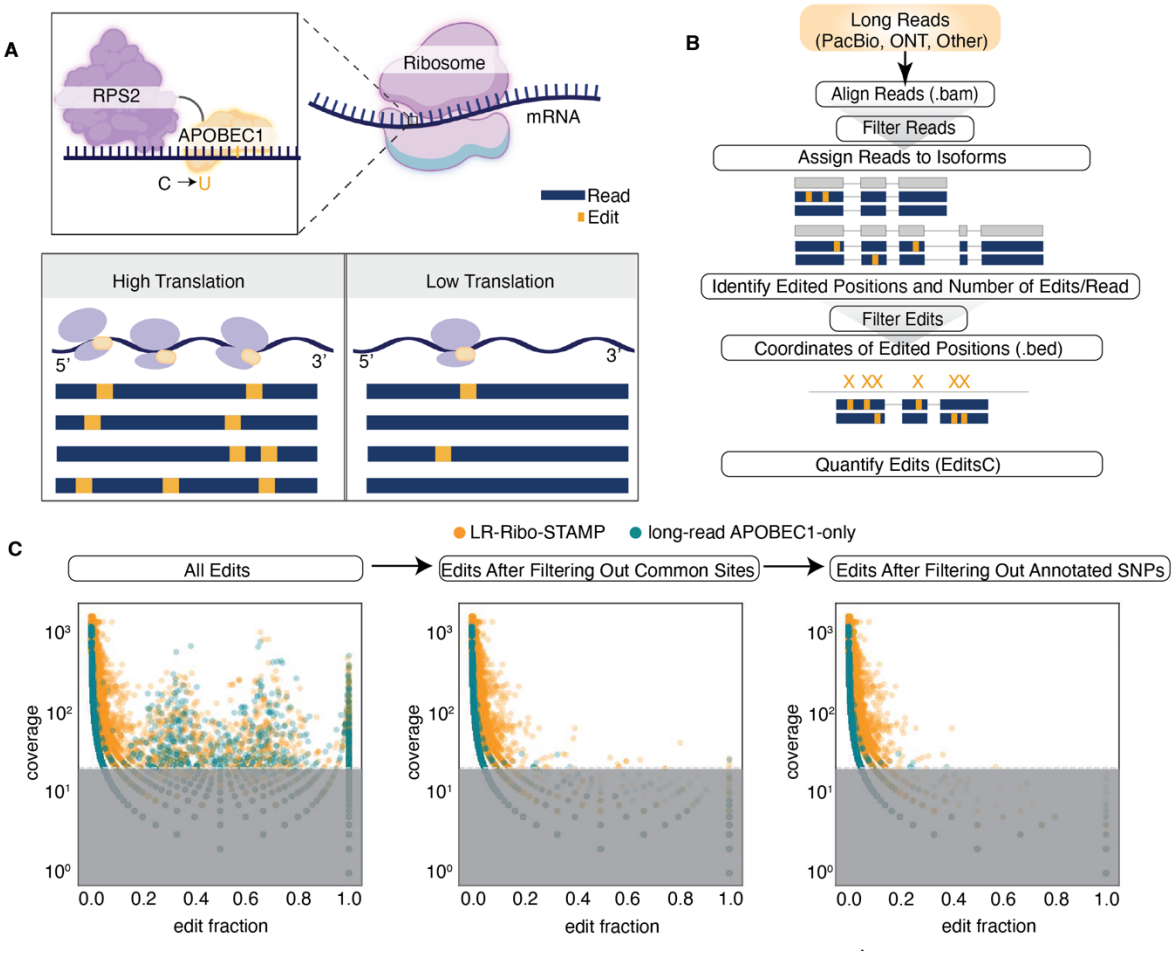


**Figure 1: Experimental and computational methods for Long-read Ribo-STAMP. (A)** Overview of the LR-Ribo-STAMP experimental system. RPS2 is fused to APOBEC1 to induce cytosine-to-uracil nucleotide edits proximal to ribosome-RNA interaction sites. More edits indicate higher translation and fewer edits 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 less than 20 reads filtered out.

## Long-Read 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-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 doxycycline for 72 hours 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 LR-Ribo-STAMP, leading to larger average EditsCs in LR-Ribo-STAMP across replicates (**Supplemental Figure 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 which 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, or 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 datasets ($R=4.09 \times 10^{-1}$, $P=2.36 \times 10^{-48}$) (**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 Figure 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). While this background may be less of an issue when comparing editing of the same gene or isoform across conditions where 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 Figure 1I)**. These results suggest that the linear model effectively identifies genes with strong signals. Overall, LR-Ribo-STAMP demonstrates positive substantial concordance with previously benchmarked short-read Ribo-STAMP data and can effectively profile gene-level translation quantification and potentially protein production in unperturbed cells.

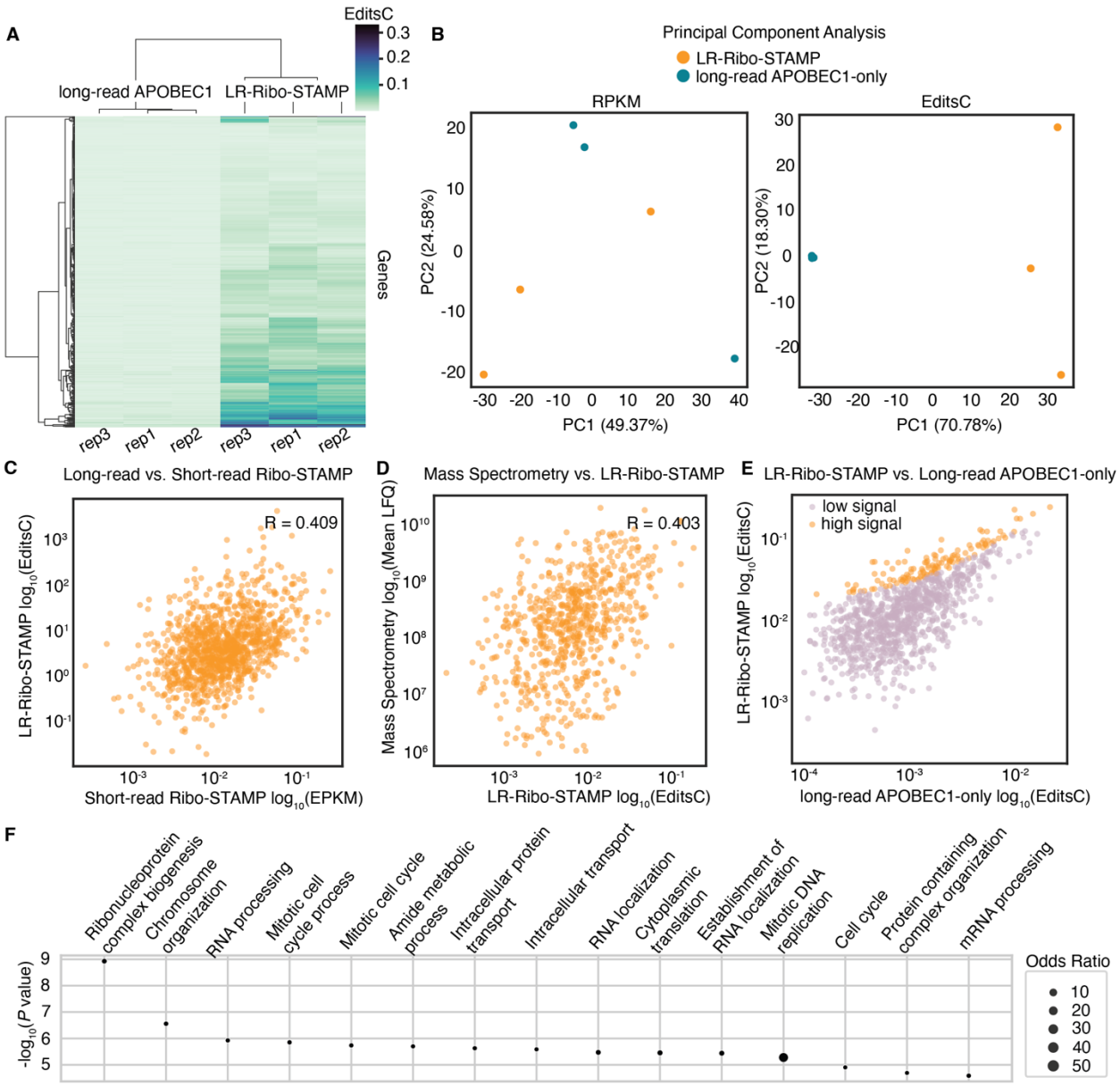


**Figure 2: Long-read Ribo-STAMP profiles translation at the gene level with good concordance with short-read data. (A)** Signal detection (EditsC) across replicates of LR-Ribo-STAMP and long-read APOBEC1-only samples at the gene level. **(B)** PCA plot showing clustering of LR-Ribo-STAMP (gold) and 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* value).

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 long-read APOBEC1-only samples (**Supplemental Figure 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 Figure 2B**). This aligns with NMD's association with transcript degradation 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 amongst 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 absence 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 dataset. However, this proved impractical due to the small number of isoforms and less differentiated EditsC profiles (**Supplemental Figure 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).

In the 5'UTR, we examined differences in length, GC content, and upstream open reading frame (uORF) prevalence between high and low translation isoforms. While the 5'UTR lengths did not significantly differ (Wilcoxon Rank Sum, $P$=4.11 x $10^{-1}$), low translation isoforms had statistically significantly higher GC content (Wilcoxon Rank Sum, $P$=7 x $10^{-3}$) (**Fig 3C-D**). Using TISdb (Wan & Qian, 2014) to obtain predicted uORFs, we found that a statistically significantly higher proportion of low translation isoforms contained predicted uORFs compared to high translation isoforms (Chi-squared test, $P$=1.5 x $10^{-3}$) (**Fig 3E**). These observations largely agree with canonical translation models (Calvo et al., 2009; Leppek et al., 2018; Pelletier & Sonenberg, 1985). 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 x $10^{-9}$) (**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 x $10^{-17}$) 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 x $10^{-6}$) and RBFOX2 (Chi-squared test, $P$=5 x $10^{-6}$) motifs were statistically significantly more prevalent in lowly translated isoforms (**Fig 3H, Supplemental Figure 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.

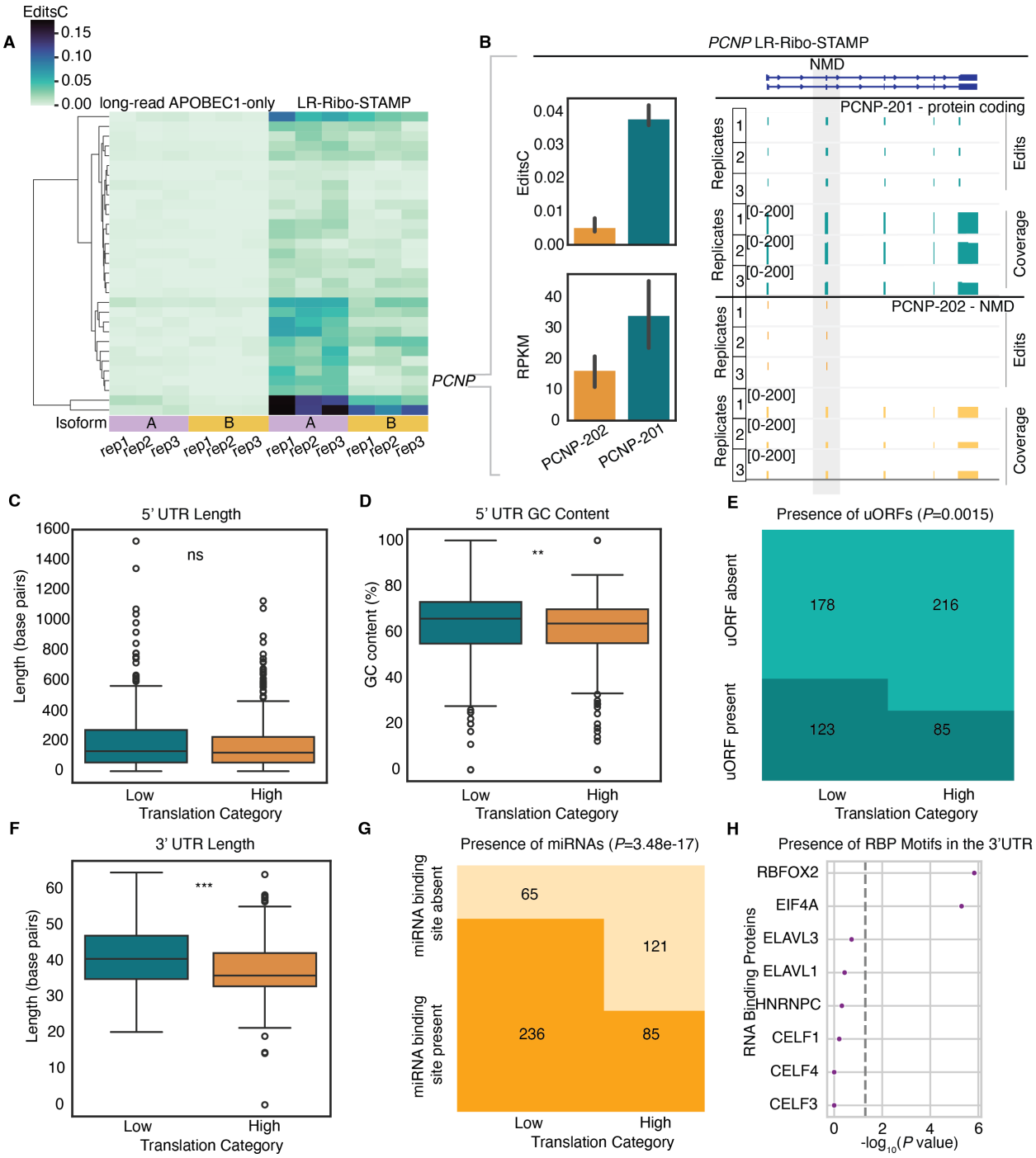


**Figure 3: Long-read Ribo-STAMP can profile translation at the mRNA isoform level for cells in an unperturbed state. (A)** Heatmap showing EditsC quantified in long-read APOBEC1-only and LR-Ribo-STAMP samples for two mRNA isoforms of the same gene. **(B)** LR-Ribo-STAMP EditsC and mRNA isoform expression 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 (G) miRNA binding sites and (H) RBP motifs in the 3'UTR sequence between highly versus lowly translated isoforms. Significance is denoted as: *** for $P \leq 0.001$, ** for $P \leq 0.01$, * for $P \leq 0.05$.

---

**Long-read 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 hours of treatment with $CoCl_2$, a commonly used hypoxia mimetic that blocks degradation of HIF1A, a transcription factor that regulates hypoxia-inducible genes (Masoud & 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 6,242 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 (Schmidt et al., 2009) (**Fig 4A**), with equal loading (**Supplemental Figure 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 co-regulation of transcription and translation (**Supplemental Figure 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 & Chánez-Cárdenas, 2019). While 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 Figure 3D-E**). Gene Ontology enrichment analysis on translationally upregulated and downregulated genes showed the enrichment of categories associated with cellular adaptation to oxygen-depletion conditions (Adzigbli et al., 2022; Lee et al., 2020; 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 where 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 (Melanson et al., 2017; Uniacke et al., 2012). 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 x $10^{-2}$) to its homolog *EIF4E2* (two-sample *t*-test, $P$=3 x $10^{-3}$) following hypoxia (**Fig 4C**). We also identified that the *CISD1* gene which harbored no mRNA expression changes showed increased translation (two-sample *t*-test, $P$=3 x $10^{-3}$), and increased TE differences (**Fig 4D, Supplemental Figure 3F**). *CISD1* has previously been shown to promote the proliferation of cancer cells, associated with poor survival, and suggested as a prognostic for breast cancer (F. Liu et al., 2022; Mittler et al., 2019; Sohn et al., 2013). Overall, LR-Ribo-STAMP effectively profiles differential translation at the gene level revealing critical shifts in regulatory molecules.

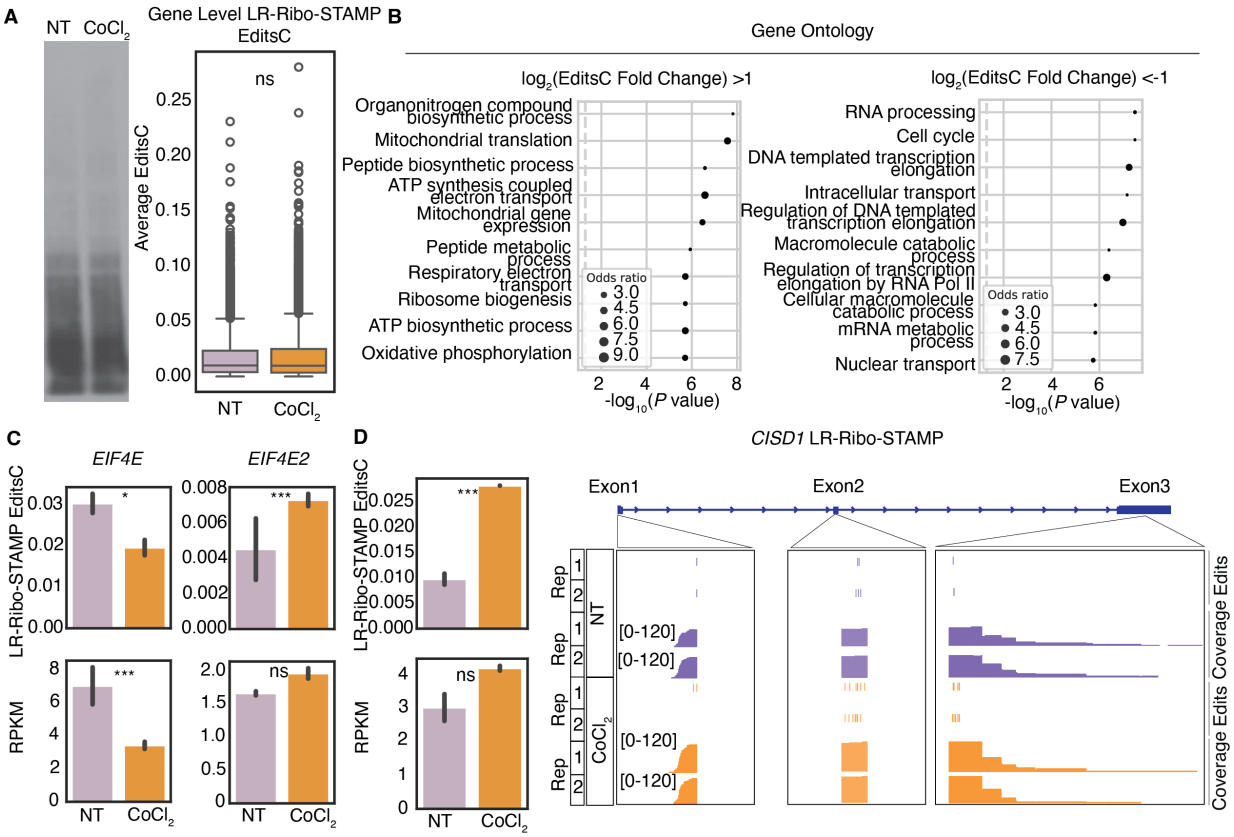


**Figure 4: Long-read Ribo-STAMP can profile changes in translation at the gene level for cells in disease state. (A)** SUnSET assay and global quantification of LR-Ribo-STAMP EditsC at the gene level for normoxia (NT) and hypoxia ($CoCl_2$) treatment conditions. **(B)** Gene Ontology analyses of genes 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 is denoted as: *** for $P \leq 0.001$, ** for $P \leq 0.01$, * for $P \leq 0.05$.

## Long-read Ribo-STAMP assesses changes in transcription and translation at mRNA isoform resolution in hypoxic conditions

Given that translatome 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 5,173 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 Figure 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, while 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 adaption 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 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 promote disease progression by increasing invasive and metastatic behavior in tumor cells (Moon et al., 2021) (**Supplemental Figure 4B**). 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 (Qiu et al., 2022) (**Supplemental Figure 4C**).

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. While 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*, which 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 alternative splicing (AS), intending to connect changes in the transcriptome to changes in the translatome. AS is a mechanism centrally placed between transcription and translation and can determine transcriptome and translatome complexity through the inclusion or exclusion of exons and introns. AS analysis using the LR-Ribo-STAMP data revealed an 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 HIF-inducible factor 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 Figure 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 48h 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 & 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 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.

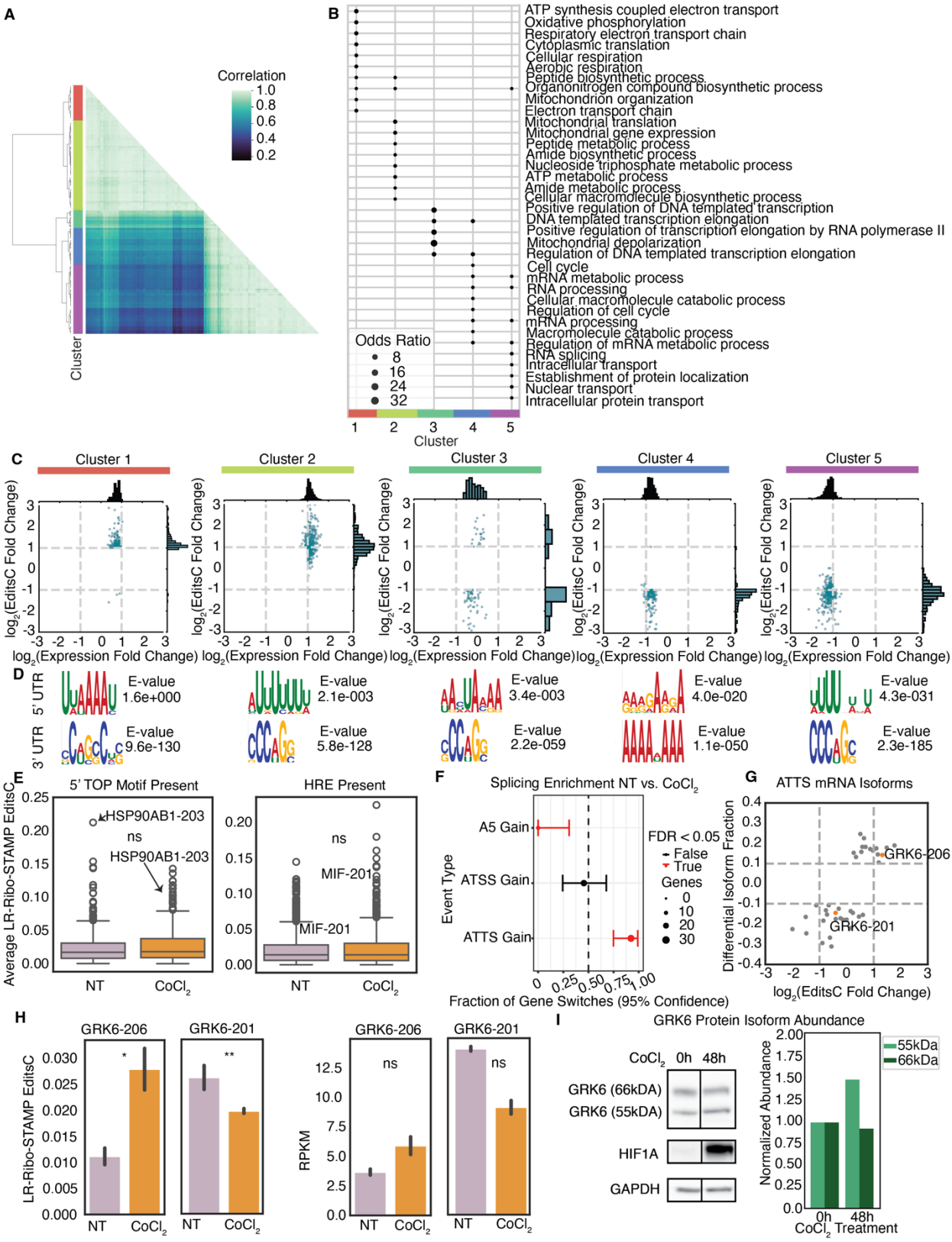


**Figure 5: Long-read Ribo-STAMP can profile changes in translation at the mRNA isoform level for cells in disease state. (A)** Hierarchical clustering of mRNA isoforms based on LR-Ribo-STAMP EditsC and RPKM metrics. The color bar indicates correlations, and the annotations

indicate clusters. **(B)** Gene Ontology enrichment analysis by cluster. **(C)** Cluster-specific changes in isoform translation ($log_2$(EditsC fold change)) versus change in expression ($log_2$(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)** Long-read 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 0h and 48h of hypoxia. Values are normalized against the 0h timepoint. Significance is denoted as: *** for $P \leq 0.001$, ** for $P \leq 0.01$, * for $P \leq 0.05$.

## 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 due 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 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 (**Fig 2-3**). When applied to evaluate differences in normoxia versus hypoxia states, 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 (**Fig 4-5**).

Our method represents a notable advance in the field by enabling quantification of translation at both 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 Technology's 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*-based method, 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 datasets and expanded isoform annotations, LR-Ribo-STAMP will be increasingly influential for profiling the translatome and transcriptome complexity. This includes the ability to analyze rare and unannotated transcripts. The extensive data generated by this method is ideal for gleaning critical regulatory pathways and mechanisms and constructing context-specific translatome 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 doxycycline (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 1ug/mL dox for 72h. Total RNA was isolated from technical triplicate 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-hour 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 hours before treatment with Puromycin (2mg/ml). Following 48 hours of Puromycin selection, cells were sorted for the top 10% of mRuby-Ribo-STAMP expressing cells on a BD Influx Cell Sorter. Cells with doxycycline-inducible Ribo-STAMP were expanded and then cultured in DMEM + 10% FBS (Gibco) containing 1ug/ml doxycycline to induce Ribo-STAMP expression and 100 uM cobalt (II) chloride (Sigma, 15862-1ML-F) to simulate hypoxia. Following 48 hours of DOX and cobalt (II) chloride treatment, cells were harvested from normoxia and induced hypoxia conditions. RNA was isolated for technical 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 2 8M SMRT cells with a 30-hour movie time on the PacBio Sequel IIe.

**SUnSET Assay**

MDA-MB-231 cells were treated with 100uM cobalt (II) chloride to induce hypoxia for 48 hours. Cells were then treated with 10ug/ml puromycin for 10 minutes and then subsequently processed for western blot analysis. To process the samples for western blot, cells were lysed in RIPA buffer (Sigma) with 200x 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 PVDF membrane. The

membrane was first incubated in Ponceau stain to obtain total protein staining for SUnSET assay normalization. Then, the membrane was blocked in 5% nonfat milk in TBST for 30 minutes and incubated overnight at 4°C with the Mouse anti-Puromycin (clone 12D10, Millipore Sigma, MABE343) antibody. The membrane was washed 3 times for 5-minutes each time in TBST, incubated for 1 hour at room temperature in 5% nonfat milk in TBST with a horse radish peroxidase-conjugated anti-mouse secondary antibody (cell signaling, 7076), and was washed 3 times for 5-minutes 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 100uM cobalt chloride II to simulate hypoxia for 24 or 48 hrs 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 polyvinylidene fluoride (PVDF) membrane. Membranes were blocked in Tris-Buffered Saline containing Tween 20 (TBST) with 5% milk for 20 minutes and probed overnight at 4C with primary antibody (Rabbit pAB anti GRK6 (N terminal) Abcam Cat# ab244364, Rabbit mAB anti HIF-1a Cell Signaling Cat# 14179, Mouse mAB anti GAPDH Millipore Cat# MAB374). Membranes were washed 3 times for 5 minutes with TBST and then probed for 1 hour at room temp in TBST containing 5% milk with secondary antibody (Anti mouse IgG, HRP linked Cell Signaling Cat# 7076, Anti rabbit IgG, HRP linked Cell Signaling Cat# 7074) diluted 1:5000. Membranes were washed 3 times for 5 minutes 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 counts 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 non-concatamer reads were generated using lima v2.9.0 with parameters: --isoseq. Reads were then refined with Iso-Seq v4.0.0's refine tool with parameters: --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 parameters: --preset ISOSEQ. HEK293XT samples were aligned to GRCh37 to maintain consistency and comparability with analyses completed by Brannan et al. (2021). GRCh38 was used 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 & Rademakers, 2023) with parameters: --raw and --tsv_stats. Reads with a quality score below 20, as assessed by NanoPlot, along with unmapped reads, supplementary 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 counts matrices and read assignments were obtained using IsoQuant v3.3.0 (Prjibelski et al., 2023) with 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 assignments (read_assignment.tsv) output from IsoQuant were used to assign individual mapped reads to isoforms. Counts matrices were used to calculate reads per kilobase of transcript per million mapped reads (RPKM). The RPKM was calculated in accordance with the following equation: RPKM = Number of Reads / (Gene Length/1000 * Total Reads/1000000). Mapped reads were determined using SAMtools v1.16 view (Danecek et al., 2021) with parameters: --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 multi-processing, 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' untranslated region, 3' untranslated region, or the coding sequence (CDS). The coordinates of these regions were determined using the GENCODE comprehensive annotations for GRCh37 (v19) and GRCh38 (v38) for 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 edit fraction = edited reads/total reads. EditsC represents the proportion of cytosines in a gene or mRNA isoform that undergo 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 & Hall, 2010) v2.29.2 merge with 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 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 translational efficiency (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, while 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 gene and mRNA isoform levels was performed using statsmodels v0.14.0, 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 was acquired from GSE155729. To compare LR-Ribo-STAMP EditsC vales with mass spectrometry data, we obtained label-free quantification (LFQ) intensity values from the LR-Ribo-STAMPProteomXchange Consortium and dataset 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 and 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 & 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 $|(\log_2(\text{EditsC hypoxia/EditsC normoxia})| \geq 1$. 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 $|\log_2(\text{expression hypoxia/expression normoxia})| \geq 1$. 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 & 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. Presence or absence of sequence elements such as 5'TOP motif and HRE were 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 & Sandelin, 2019).

## 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.

## SOFTWARE AVAILABILITY

Source code and analysis scripts for edit quantification are available at https://github.com/YeoLab/LR-Ribo-STAMP and in Supplemental Code.

## COMPETING INTEREST STATEMENT

G.W.Y. is a co-founder, member of the Board of Directors, on the SAB, equity holder, and paid consultant for Locanabio (until 31 December 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.

## ACKNOWLEDGEMENTS

This work was supported by NIH grants HG004659, HG011864, HG009889 and MH126719 to G.W.Y. P.J. was supported by NIH institutional training grant T32 GM145427 and ARCS Scholarship. K.W.B. is supported by NIH/NINDS K22 NS112678, NIH/NCI R01 CA284315 and CPRIT Award RR220017. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing a Sequel II. The authors would like to thank members of the Yeo Lab for helpful discussions and suggestion during the preparation of this manuscript. 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 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 was 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, E. P., Wimmers, K., Sokolova, I. M., & Ponsuksili, S. (2022). Effects of hypoxia and reoxygenation on mitochondrial functions and transcriptional profiles of isolated brain and muscle porcine cells. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-24386-0

Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, *4*, e05005. https://doi.org/10.7554/eLife.05005

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 30. https://doi.org/10.1186/s13059-020-1935-5

Arsham, A. M., Howell, J. J., & Simon, M. C. (2003). A novel hypoxia-inducible factor-independent hypoxic response regulating mammalian target of rapamycin and its targets. *The Journal of Biological Chemistry*, *278*(32), 29655–29660. https://doi.org/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, C. H., Ramirez Flores, R. O., & Saez-Rodriguez, J. (2022). decoupleR: Ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, *2*(1), vbac016. https://doi.org/10.1093/bioadv/vbac016

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, *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. *Biomedicine & Pharmacotherapy* =

*Biomedecine & Pharmacotherapie*, *57*(8), 333–340. https://doi.org/10.1016/s0753-3322(03)00098-2

Brannan, K. W., Chaim, I. A., Marina, R. J., Yee, B. A., Kofman, E. R., Lorenz, D. A., Jagannatha, P., Dong, K. D., Madrigal, A. A., Underwood, J. G., & Yeo, G. W. (2021). Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. *Nature Methods*, *18*(5), Article 5. https://doi.org/10.1038/s41592-021-01128-0

Calvo, S. E., Pagliarini, D. J., & Mootha, V. K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7507–7512. https://doi.org/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*(24), 678. https://doi.org/10.1186/s12859-019-3244-0

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

De Coster, W., & Rademakers, R. (2023). NanoPack2: Population-scale evaluation of long-read sequencing data. *Bioinformatics*, *39*(5), btad311. https://doi.org/10.1093/bioinformatics/btad311

Dorsam, R. T., & Gutkind, J. S. (2007). G-protein-coupled receptors and cancer. *Nature Reviews Cancer*, *7*(2), Article 2. https://doi.org/10.1038/nrc2069

Epstein, A. C., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., & Ratcliffe, P. J. (2001). C. elegans EGL-9 and mammalian homologs define a family

of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*, *107*(1), 43–54. https://doi.org/10.1016/s0092-8674(01)00507-4

Flamand, M. N., Ke, K., Tamming, R., & Meyer, K. D. (2022). Single-molecule identification of the target RNAs of different RNA binding proteins simultaneously in cells. *Genes & Development*, *36*(17–18), 1002–1015. https://doi.org/10.1101/gad.349983.122

Floor, S. N., & Doudna, J. A. (2016). Tunable protein synthesis by transcript isoforms in human cells. *eLife*, *5*, e10921. https://doi.org/10.7554/eLife.10921

Foord, C., Hsu, J., Jarroux, J., Hu, W., Belchikov, N., Pollard, S., He, Y., Joglekar, A., & Tilgner, H. U. (2023). The variables on RNA molecules: Concert or cacophony? Answers in long-read sequencing. *Nature Methods*, *20*(1), Article 1. https://doi.org/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*(23), 3060–3066. https://doi.org/10.1093/bioinformatics/btt537

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biology*, *8*(2), R24. https://doi.org/10.1186/gb-2007-8-2-r24

Harris, A. L. (2002). Hypoxia—A key regulatory factor in tumour growth. *Nature Reviews Cancer*, *2*(1), Article 1. https://doi.org/10.1038/nrc704

Hegazi, S., Cheng, A. H., Krupp, J. J., Tasaki, T., Liu, J., Szulc, D. A., Ling, H. H., Rios Garcia, J., Seecharran, S., Basiri, T., Amiri, M., Anwar, Z., Ahmad, S., Nayal, K., Sonenberg, N., Liu, B.-H., Cheng, H.-L. M., Levine, J. D., & Cheng, H.-Y. M. (2022). UBR4/POE facilitates secretory trafficking to maintain circadian clock synchrony. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-29244-1

633 Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016). Translational control by 5′-
634 untranslated regions of eukaryotic mRNAs. *Science*, *352*(6292), 1413–1416.
635 https://doi.org/10.1126/science.aad9868
636 Hu, Y., Jiang, Z., Chen, K., Zhou, Z., Zhou, X., Wang, Y., Yang, J., Zhang, B., Wen, L., & Tang,
637 F. (2023). scNanoATAC-seq: A long-read single-cell ATAC sequencing method to detect
638 chromatin accessibility and genetic variants simultaneously within an individual cell. *Cell*
639 *Research*, *33*(1), Article 1. https://doi.org/10.1038/s41422-022-00730-x
640 Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-Wide
641 Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling.
642 *Science*, *324*(5924), 218–223. https://doi.org/10.1126/science.1168978
643 Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., & Muñoz, M. J. (2013).
644 Alternative splicing: A pivotal step between eukaryotic transcription and translation.
645 *Nature Reviews Molecular Cell Biology*, *14*(3), Article 3. https://doi.org/10.1038/nrm3525
646 Lee, P., Chandel, N. S., & Simon, M. C. (2020). Cellular adaptation to hypoxia through hypoxia
647 inducible factors and beyond. *Nature Reviews Molecular Cell Biology*, *21*(5), Article 5.
648 https://doi.org/10.1038/s41580-020-0227-y
649 Leppek, K., Das, R., & Barna, M. (2018). Functional 5′ UTR mRNA structures in eukaryotic
650 translation regulation and how to find them. *Nature Reviews Molecular Cell Biology*,
651 *19*(3), Article 3. https://doi.org/10.1038/nrm.2017.103
652 Li, J. J., Chew, G.-L., & Biggin, M. D. (2019). Quantitative principles of cis-translational control
653 by general mRNA sequence features in eukaryotes. *Genome Biology*, *20*(1), 162.
654 https://doi.org/10.1186/s13059-019-1761-9
655 Lin, Y., Kwok, S., Hein, A. E., Thai, B. Q., Alabi, Y., Ostrowski, M. S., Wu, K., & Floor, S. N.
656 (2023). RNA molecular recording with an engineered RNA deaminase. *Nature Methods*,
657 *20*(12), Article 12. https://doi.org/10.1038/s41592-023-02046-z

Liu, F., Dong, Y., Zhong, F., Guo, H., & Dong, P. (2022). CISD1 Is a Breast Cancer Prognostic Biomarker Associated with Diabetes Mellitus. *Biomolecules*, *13*(1), 37. https://doi.org/10.3390/biom13010037

Liu, Z., Quinones-Valdez, G., Fu, T., Huang, E., Choudhury, M., Reese, F., Mortazavi, A., & Xiao, X. (2023). L-GIREMI uncovers RNA editing sites in long-read RNA-seq. *Genome Biology*, *24*(1), 171. https://doi.org/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., Zhou, Y., Lin, P., Wu, B., Yuan, Y., Zhao, J., Xu, W., & Zhao, S. (2024). Hypoxia induces mitochondrial protein lactylation to limit oxidative phosphorylation. *Cell Research*, *34*(1), Article 1. https://doi.org/10.1038/s41422-023-00864-6

Marx, V. (2023). Method of the year: Long-read sequencing. *Nature Methods*, *20*(1), Article 1. https://doi.org/10.1038/s41592-022-01730-w

Masoud, G. N., & Li, W. (2015). HIF-1α pathway: Role, regulation and intervention for cancer therapy. *Acta Pharmaceutica Sinica B*, *5*(5), 378–389. https://doi.org/10.1016/j.apsb.2015.05.007

Medina-Munoz, H. C., Kofman, E., Jagannatha, P., Boyle, E. A., Yu, T., Jones, K. L., Mueller, J. R., Lykins, G. D., Doudna, A. T., Park, S. S., Blue, S. M., Ranzau, B. L., Kohli, R. M., Komor, A. C., & Yeo, G. W. (2024). Expanded palette of RNA base editors for comprehensive RBP-RNA interactome studies. *Nature Communications*, *15*(1), 875. https://doi.org/10.1038/s41467-024-45009-4

Melanson, G., Timpano, S., & Uniacke, J. (2017). The eIF4E2-Directed Hypoxic Cap-Dependent Translation Machinery Reveals Novel Therapeutic Potential for Cancer Treatment. *Oxidative Medicine and Cellular Longevity*, *2017*, e6098107. https://doi.org/10.1155/2017/6098107

682 Mitschka, S., & Mayr, C. (2022). Context-specific regulation and function of mRNA alternative
683 polyadenylation. *Nature Reviews Molecular Cell Biology*, *23*(12), Article 12.
684 https://doi.org/10.1038/s41580-022-00507-5
685 Mittler, R., Darash-Yahana, M., Sohn, Y. S., Bai, F., Song, L., Cabantchik, I. Z., Jennings, P. A.,
686 Onuchic, J. N., & Nechushtai, R. (2019). NEET Proteins: A New Link Between Iron
687 Metabolism, Reactive Oxygen Species, and Cancer. *Antioxidants & Redox Signaling*,
688 *30*(8), 1083–1095. https://doi.org/10.1089/ars.2018.7502
689 Moon, E. J., Mello, S. S., Li, C. G., Chi, J.-T., Thakkar, K., Kirkland, J. G., Lagory, E. L., Lee, I.
690 J., Diep, A. N., Miao, Y., Rafat, M., Vilalta, M., Castellini, L., Krieg, A. J., Graves, E. E.,
691 Attardi, L. D., & Giaccia, A. J. (2021). The HIF target MAFF promotes tumor invasion
692 and metastasis through IL11 and STAT3 signaling. *Nature Communications*, *12*(1),
693 Article 1. https://doi.org/10.1038/s41467-021-24631-6
694 Muñoz-Sánchez, J., & Chánez-Cárdenas, M. E. (2019). The use of cobalt chloride as a
695 chemical hypoxia model. *Journal of Applied Toxicology*, *39*(4), 556–570.
696 https://doi.org/10.1002/jat.3749
697 Nickless, A., Bailis, J. M., & You, Z. (2017). Control of gene expression through the nonsense-
698 mediated RNA decay pathway. *Cell & Bioscience*, *7*(1), 26.
699 https://doi.org/10.1186/s13578-017-0153-7
700 Oliveto, S., Mancino, M., Manfrini, N., & Biffo, S. (2017). Role of microRNAs in translation
701 regulation and cancer. *World Journal of Biological Chemistry*, *8*(1), 45–56.
702 https://doi.org/10.4331/wjbc.v8.i1.45
703 Pearce, L. R., Komander, D., & Alessi, D. R. (2010). The nuts and bolts of AGC protein kinases.
704 *Nature Reviews Molecular Cell Biology*, *11*(1), Article 1. https://doi.org/10.1038/nrm2822
705 Pelletier, J., & Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure
706 within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency.
707 *Cell*, *40*(3), 515–526. https://doi.org/10.1016/0092-8674(85)90200-4

Prjibelski, A. D., Mikheenko, A., Joglekar, A., Smetanin, A., Jarroux, J., Lapidus, A. L., & Tilgner, H. U. (2023). Accurate isoform discovery with IsoQuant using long reads. *Nature Biotechnology*, *41*(7), Article 7. https://doi.org/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., Deng, J., & Wu, Y. (2022). EBF1 promotes triple-negative breast cancer progression by surveillance of the HIF1α pathway. *Proceedings of the National Academy of Sciences*, *119*(28), e2119518119. https://doi.org/10.1073/pnas.2119518119

Quattrone, A., & Dassi, E. (2019). The Architecture of the Human RNA-Binding Protein Regulatory Network. *iScience*, *21*, 706–719. https://doi.org/10.1016/j.isci.2019.10.058

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org

Reixachs-Solé, M., Ruiz-Orera, J., Albà, M. M., & Eyras, E. (2020). Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15634-w

Riley, T. R., Slattery, M., Abe, N., Rastogi, C., Liu, D., Mann, R. S., & Bussemaker, H. J. (2014). SELEX-seq: A method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods in Molecular Biology (Clifton, N.J.)*, *1196*, 255–278. https://doi.org/10.1007/978-1-4939-1242-1_16

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), Article 1. https://doi.org/10.1038/nbt.1754

Schmidt, E. K., Clavarino, G., Ceppi, M., & Pierre, P. (2009). SUnSET, a nonradioactive method to monitor protein synthesis. *Nature Methods*, *6*(4), 275–277. https://doi.org/10.1038/nmeth.1314

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311. https://doi.org/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. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-55296-3

Simpson, B. P., Yrigollen, C. M., Izda, A., & Davidson, B. L. (2023). Targeted long-read sequencing captures CRISPR editing and AAV integration outcomes in brain. *Molecular Therapy*, *31*(3), 760–773. https://doi.org/10.1016/j.ymthe.2023.01.004

Sohn, Y.-S., Tamir, S., Song, L., Michaeli, D., Matouk, I., Conlan, A. R., Harir, Y., Holt, S. H., Shulaev, V., Paddock, M. L., Hochberg, A., Cabanchick, I. Z., Onuchic, J. N., Jennings, P. A., Nechushtai, R., & Mittler, R. (2013). NAF-1 and mitoNEET are central to human breast cancer proliferation by maintaining mitochondrial homeostasis and promoting tumor growth. *Proceedings of the National Academy of Sciences*, *110*(36), 14676–14681. https://doi.org/10.1073/pnas.1313198110

Sterne-Weiler, T., Martinez-Nunez, R. T., Howard, J. M., Cvitovik, I., Katzman, S., Tariq, M. A., Pourmand, N., & Sanford, J. R. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Research*, *23*(10), 1615–1623. https://doi.org/10.1101/gr.148585.112

Thoreen, C. C., Chantranupong, L., Keys, H. R., Wang, T., Gray, N. S., & Sabatini, D. M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, *485*(7396), Article 7396. https://doi.org/10.1038/nature11083

Tripathi, V. K., Subramaniyan, S. A., & Hwang, I. (2019). Molecular and Cellular Response of Co-cultured Cells toward Cobalt Chloride (CoCl2)-Induced Hypoxia. *ACS Omega*, *4*(25), 20882–20893. https://doi.org/10.1021/acsomega.9b01474

Uniacke, J., Holterman, C. E., Lachance, G., Franovic, A., Jacob, M. D., Fabian, M. R., Payette, J., Holcik, M., Pause, A., & Lee, S. (2012). An oxygen-regulated switch in the protein synthesis machinery. *Nature*, *486*(7401), Article 7401. https://doi.org/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*(21), 4469–4471. https://doi.org/10.1093/bioinformatics/btz247

Wan, J., & Qian, S.-B. (2014). TISdb: A database for alternative translation initiation in mammalian cells. *Nucleic Acids Research*, *42*(D1), D845–D850. https://doi.org/10.1093/nar/gkt1085

Yao, S., Ertay, A., Zhou, Y., Yao, L., Hill, C., Chen, J., Guan, Y., Sun, H., Ewing, R. M., Liu, Y., Lv, X., & Wang, Y. (2021). GRK6 Depletion Induces HIF Activity in Lung Adenocarcinoma. *Frontiers in Oncology*, *11*, 654812. https://doi.org/10.3389/fonc.2021.654812

Zarrilli, G., Businello, G., Dieci, M. V., Paccagnella, S., Carraro, V., Cappellesso, R., Miglietta, F., Griguolo, G., Guarneri, V., Lo Mele, M., & Fassan, M. (2020). The Tumor Microenvironment of Primitive and Metastatic Breast Cancer: Implications for Novel Therapeutic Strategies. *International Journal of Molecular Sciences*, *21*(21), 8102. https://doi.org/10.3390/ijms21218102

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

Pratibha Jagannatha, Alexandra T Tankka, Daniel A Lorenz, et al.

*Genome Res.* published online June 21, 2024
Access the most recent version at doi:10.1101/gr.279176.124