NAR Genomics and Bioinformatics, 2026, **8**, lqag130
https://doi.org/10.1093/nargab/lqag130
Computational Methods for Transcriptome Analysis

Rp-Bp and the Ribotools box for translatome profiling

Etienne Boileau[1,2,3], Maja Bencun[1,2,3], Francesca Tuorto[4,5], Isabel S. Naarmann-de Vries[1,2,3], Pauline Fahjen[6], Philipp Mertins[6,7], Christoph Dieterich[1,2,3,*]

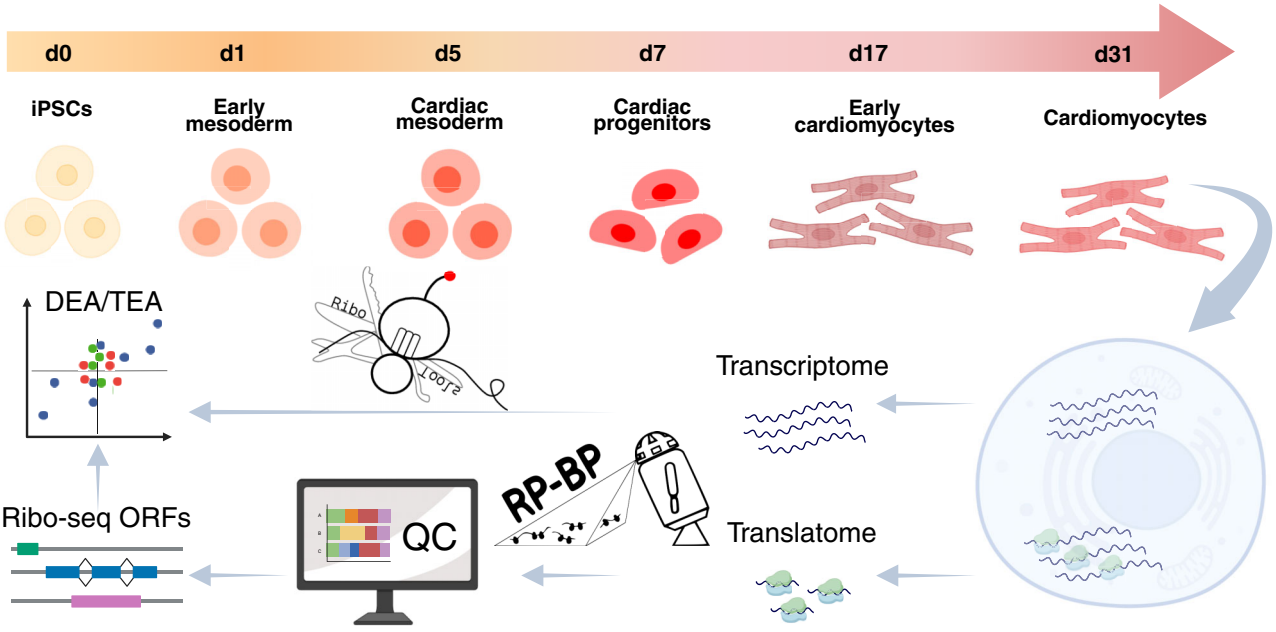
[1]Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Im Neuenheimer Feld 669, 69120 Heidelberg, Germany
[2]Klaus Tschira Institute for Integrative Computational Cardiology, University of Heidelberg, Im Neuenheimer Feld 669, 69120 Heidelberg, Germany
[3]German Center for Cardiovascular Research (DZHK) – Partner site Heidelberg/Mannheim, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany
[4]Division of Biochemistry, Mannheim Institute for Innate Immunoscience (MI3), Mannheim Cancer Center, Medical Faculty Mannheim, Heidelberg University, Franz-Volhard-Str. 6, 68167 Mannheim, Germany
[5]Center for Molecular Biology of Heidelberg University (ZMBH), DKFZ-ZMBH Alliance, 69120 Heidelberg, Germany
[6]Max Delbrück Center for Molecular Medicine (MDC), Robert-Rössle-Straße 10, 13125 Berlin, Germany
[7]Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

*To whom correspondence should be addressed. Email: christoph.dieterich@uni-heidelberg.de

Abstract

We present the Ribotools "box" for the analysis of matched ribosome profiling (Ribo-seq) and RNA sequencing data. Ribotools can be used in standalone mode or in combination with Rp-Bp for Ribo-seq quality control, open reading frames discovery, translation efficiency, and differential expression analyses. In this work, we use Ribotools to profile the differentiation of human induced pluripotent stem cells (hiPSCs) into cardiomyocytes to reveal a remarkably coherent transition from plastic, signaling-driven to stable, structurally committed cardiac cells, staged through multiple layers of translational control. Together with complementary proteomics, the resulting dataset provides a rich multi-omics resource for investigating the hiPSC model of cardiac lineage commitment. Additionally, we present a reproducible and extensible pipeline for benchmarking translation efficiency analysis tools. Ribotools and Rp-Bp are available on Bioconda and help to make research accessible and reproducible. Detailed documentation, test datasets, and tutorials are available at https://ribotools.readthedocs.io and https://rp-bp.readthedocs.io.

Graphical abstract



Received: April 9, 2026. **Revised:** September 11, 2026. **Accepted:** September 15, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Ribosome profiling (Ribo-seq) is an RNA sequencing (RNA-seq)-based readout of RNA translation. Isolation and deep sequencing of ribosome-protected RNA fragments (RPFs, or ribosome footprints) provide a genome-wide snapshot of the translatome at sub-codon resolution. Since 2010, Ribo-seq has been widely applied to many organisms, tissues, or cells, revealing a dark translatome of thousands of unannotated open reading frames (ORFs) [1–3]. The identification, characterization, and annotation of small functional or regulatory ORFs continue to attract substantial interest [4, 5], together with efforts to understand how tissue-, stage-, or disease-specific translational regulation can be harnessed for both research and therapeutic applications [6–8]. Recent efforts have focused on developing consensus public resources, such as databases and web servers [9–13], and on defining standardized annotations and guidelines for non-canonical ORF research [14, 15]. Standalone, bundled, and online Ribo-seq-based tools have been developed for use as data resources, for ORF discovery and differential translation [16]. Most methods use the 3-nt periodic signature of translating ribosomes to identify Ribo-seq ORFs, including Rp-Bp [17]. Since its publication, Rp-Bp has been used in a number of studies where unsupervised selection of read lengths was required for robust periodicity estimation [18–22].

Although several methods have been developed specifically for translation efficiency analysis (TEA), the field currently lacks a robust, modular, end-to-end (E2E) solution that is actively maintained, well documented, and enables a principled classification of features into distinct layers of gene regulation. We present here Ribotools, a comprehensive toolbox that, in combination with Rp-Bp, provides an E2E solution for Ribo-seq quality control (QC), ORF discovery, and TEA. We provide a benchmarking pipeline for TEA tools, publicly available at https://github.com/eboileau/ribotools-benchmark, and compare Ribotools with available ORF discovery tools. We show that Ribotools stands out as the best all-around performer. We generated a time course dataset of differentiating human induced pluripotent stem cells (hiPSCs) into cardiomyocytes (CMs). Ribotools allowed us to reveal how translational regulation follows a highly dynamic trajectory, shifting from early buffering of lineage programs to translational forwarding during commitment, followed by suppression of signaling and remodeling pathways at later stages. Our results also include a comprehensive atlas of Ribo-seq ORFs, complemented by mass spectrometry, as well as interactive dashboards hosted at https://rpbp.dieterichlab.org/hIPSCCMdiff/profile_construction and https://rpbp.dieterichlab.org/hIPSCCMdiff/predictions. These data will serve as a resource and blueprint for investigating hiPSCs as a model of cardiac development.

Materials and methods

Maintenance of hiPSCs

The hiPSC line UMGi014-C clone 14 (isWT1.14) was provided by Lukas Cyganek (Stem Cell Unit, University Medical Center Göttingen). Cells were cultured in Stem MACS iPS Brew XF Medium (Miltenyi Biotec) under humidified normoxic conditions (37°C, 5% CO_2). The culture medium was changed daily. Cultures were routinely passaged at a ratio of 1:6 every 4–5 days following dissociation using Versene Solution (Thermo Fisher). Culture plates were coated with 17μg/cm^2 Growth Factor Reduced Matrigel (Corning) prior to seeding. Cells were plated in Stem MACS iPS Brew XF Medium supplemented with 2 μM Thiazovivin (Millipore) for the first 24 h after dissociation.

Differentiation of hiPSCs into ventricular cardiomyocytes

Human iPSCs were differentiated into ventricular CMs as previously described [23]. Briefly, cells were detached using Versene Solution (Thermo Fisher) and plated on 6-well culture plates pre-coated with 17 μg/cm^2 Growth Factor Reduced Matrigel (Corning) using 150 000 cells/well in a final volume of 2 ml Stem MACS iPS Brew XF Medium supplemented with 2 μM Thiazovivin (Millipore) per well. Differentiation was started (day 0) when cells reached >95% confluency by replacing iPS culture medium with 3 ml/well Cardio Differentiation Medium [RPMI 1640 with GlutaMAX and HEPES (Thermo Fisher) supplemented with 0.2 mg/ml L-ascorbic acid 2-phosphate (Sigma–Aldrich), 0.5 mg/ml human recombinant Albumin (Sigma–Aldrich), and 4 μM CHIR99021 (Merck Millipore)]. The next day, an additional 3 ml Cardio Differentiation Medium were added on top, creating a CHIR99021 concentration gradient. On day 3, the medium was switched to Cardio Differentiation Medium supplemented with 5 μM IWP2 (Merck Millipore). Day 5 and onward, the medium was changed to 3 ml Cardio Differentiation Medium every 2 days. From day 7 onward, the medium was changed to 3 ml Cardio Culture Medium [RPMI 1640 with GlutaMAX and HEPES (Thermo Fisher) supplemented with 1× B27 (Thermo Fisher)] every 2 days. On day 10, metabolic selection was performed using Cardio Selection Medium [RPMI 1640 without glucose and HEPES, with 0.2mg/ml L-ascorbic acid 2-phosphate, 0.5 mg/ml human recombinant Albumin (Sigma–Aldrich), 4 mM lactate/HEPES] for 3 days. On day 15, cells were detached using 0.25% Trypsin ethylenediaminetetraacetic acid (Thermo Fisher) and replated in 2ml Cardio Culture Medium supplemented with 20% FCS (Thermo Fisher) and 2 μM Thiazovivin on Matrigel pre-coated 6-well culture plates. A second metabolic selection was performed on day 20 as described earlier. From day 23 onward, cultures were maintained in Cardio Culture Medium with regular media changes every 2–3 days.

Sample collection

Samples were collected at days 0, 1, 5, 7, 17, and 31 post-differentiation start. For transcriptome and translatome analysis, cells were treated with cycloheximide (CHX) (100 μg/ml) for 5 min at 37°C, washed once with ice cold phosphate buffered saline (PBS)/CHX (100 μg/ml), and lysed in 600 μl of lysis buffer [20 mM Tris–HCl, pH 7.4, 5 mM $MgCl_2$, 150 mM NaCl, 1% Triton X-100, 100 μg/ml CHX, 1×cOmplete protease inhibitors (Roche)] per 6-well plate. One 6-well plate was used as one replicate. Lysates were clarified at 10 000 rpm for 10 min at 4°C, and the supernatants were snap-frozen in liquid nitrogen until all samples were collected. For mass spectrometry, cells were detached using a cell scraper into 1ml of ice-cold PBS. Cell pellets were collected by centrifugation (1200 rpm, 4°C, 10 min), washed once with ice-cold PBS, and snap frozen in liquid nitrogen.

Ribosome profiling

Snap-frozen polysome lysates were thawed on ice. Ribosome footprinting was performed as previously described in Tuorto *et al.* [24]. Briefly, 400 µl of polysome lysates were digested with 8 µl RNase I (100 U/µl) and 4 µl TURBO DNase I (1 U/µl) for 45 min at RT with occasional shaking. To stop digestion 20 µl of RNasin ribonuclease inhibitor (40 U/µl, Promega) was added to each sample. Tubes were stored on ice, and 300 µl of each sample was loaded onto a 17.5%–50% linear sucrose gradient. Samples were centrifuged at 35 000 rpm for 2.5 h at 4°C in a Beckmann SW60 rotor. Gradients were eluted using a Foxy R1 Fraction Collector (Teledyne ISCO), and profiles were recorded by continuously monitoring absorbance at 254 nm. Fractions containing the monosome peaks in the recorded profiles were used for library preparation. Ribosome-protected footprint RNA was end-repaired with T4 polynucleotide kinase (Thermo Fisher) and size-selected for 28–31 nt on a 15% polyacrylamide TBE-UREA gel. Sequencing libraries were prepared using the NEBNEXT Small RNA Library Prep kit for Illumina (NEB). Libraries were pooled based on molarity and sequenced at GeneCore (EMBL) on a NextSeq 2000 platform.

RNA sequencing

Total RNA was extracted from 100 µl of polysome lysate using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. For each differentiation time point, two biological replicates were processed. Ribosomal RNA (rRNA) was depleted using the NEBNext rRNA Depletion Kit v2 (Human/Mouse/Rat) (New England Biolabs). Sequencing libraries were prepared from 200 ng total RNA using the NEBNext UltraExpress RNA Library Prep Kit (NEB #E3330L) according to the manufacturer's instructions, with a fragmentation time of 15 min. Libraries were pooled based on molarity and sequenced at GeneCore (EMBL) on a NextSeq 2000 platform.

Genome annotation

The human assembly GRCh38 was used with Ensembl human gene annotations (Ensembl release 102) to guide a *de novo* transcriptome assembly for the differentation of hiPSCs to CMs. Paired-end RNA-seq data were trimmed for residual adapter sequences and quality-filtered using Flexbar. Bowtie2 was used for *in silico* rRNA removal. The remaining reads were mapped using STAR. The transcripts were assembled using StringTie. The annotations were merged into a non-redundant set of features and compared with the reference annotation using GffCompare. All details and references are in the Supplementary Methods S1.

Ribotools pipeline overview

The Ribotools pipeline supports multiple analysis workflows. It operates as a standalone solution to generate periodic (Ribo-seq) or trimmed (RNA-seq) alignments and read-count tables (quantification); it can be used for translation efficiency (TE/DE analyses) alone or in combination with Rp-Bp for streamlined QC (Fig. 1A). When used in conjunction with the abundance estimation pipeline, the TEA pipeline automatically generates the required input files. Alternatively, it can be run directly on count data produced by an external workflow. For analyzes based on Ribo-seq ORFs, Ribotools provides functionality to generate a dedicated GTF annotation, enabling the abundance estimation pipeline to quantify ORF-level expression and subsequently perform TE or differential expression analyses at the ORF level. The QC dashboard allows users to visualize read filtering, length distributions, periodicity estimates, metagene profiles, etc. Together with the ORF discovery pipeline, the corresponding dashboard provides multiple views to facilitate ORF discovery, including an integrated IGV browser for visual exploration of predicted Ribo-seq ORFs. The applications can be run locally on the user's browser, remotely, or served to share results with collaborators. The Ribotools and Rp-Bp pipelines are designed to support the complete analysis workflow, from raw FASTQ files to the final set of translated Ribo-seq ORFs and/or translationally regulated features. Each pipeline can also be started from any intermediate step, provided the input data are appropriately prepared and organized (Fig. 1B). A single YAML configuration file can be used for all steps. The directory structure is automatically generated and used to identify existing data on the file system. In the background, Ribotools uses HTSeq for abundance estimation and DEseq2 to model TE. Differential TE is tested in a two-step procedure using a likelihood ratio test (LRT) within the DESeq2 negative-binomial GLM framework: LRT for the global model fit, and Wald tests with fold change threshold for extracting directional contrasts. Ribotools then automatically classifies significant features into four regulatory classes, according to the terminology defined in the deltaTE protocol [25]. All details and references are in the Supplementary Methods S2.

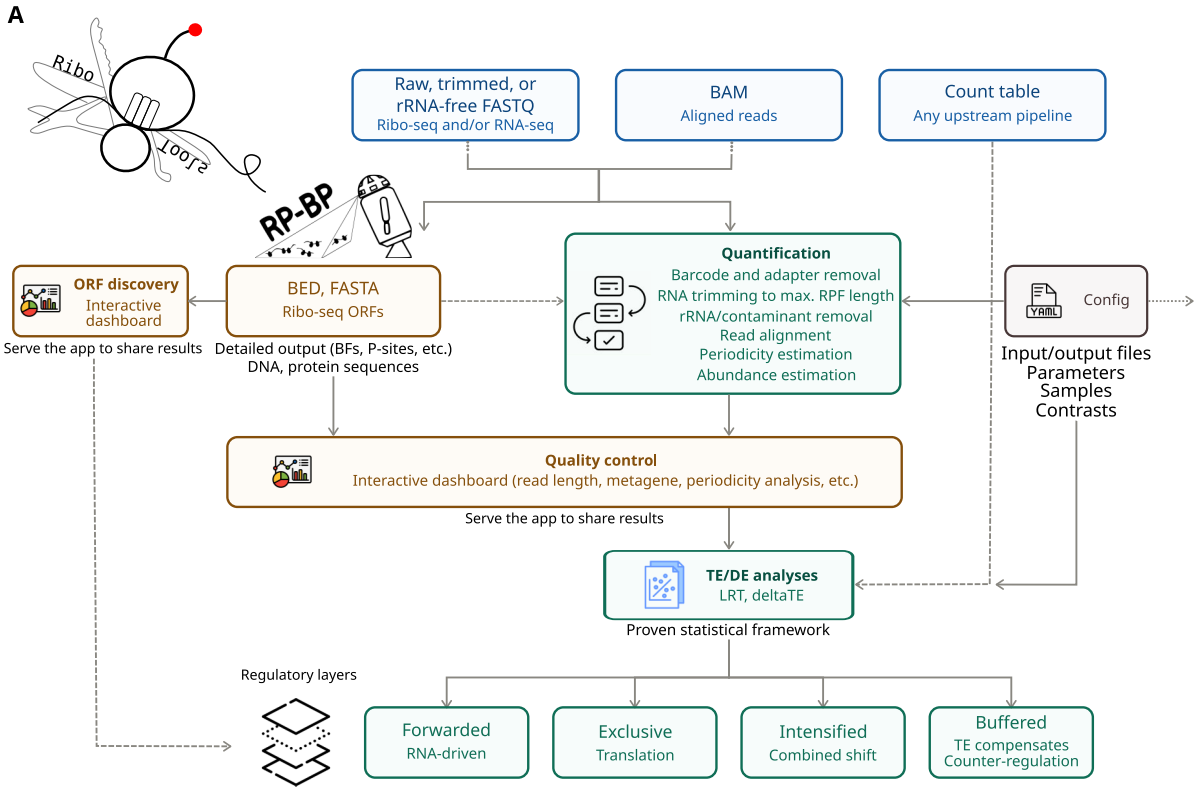
Rp-Bp pipeline overview

The pipeline consists of two modules: the ORF profile construction and the Ribo-seq ORFs prediction. The Bayesian approach was previously described [17]. In this new release, PyStan 2 was replaced with the CmdStanPy API. All models are now pre-compiled and shipped with the Bioconda package. Here, we used the newly generated transcriptome annotation to predict *de novo* translation events by directly assessing the periodicity of the Ribo-seq signal, without the need for model training or fine-tuning. Rp-Bp automatically assigns identifiers and labels to ORFs using their position relative to the annotated transcript-exon structure. The ORF identifiers include the host transcript and genomic coordinates, and the labels follow the standardized annotation of ORFs suggested by the TransCODE Consortium [14], enabling users and non-specialists to easily catalog alternative ORFs and recognize those that may be of particular interest, simplifying downstream analyzes. All details are in the Supplementary Methods S3.

Results

Ribotools stands out as the strongest all-around tool

Among tools specifically designed for TEA, Ribotools is the only fully modular pipeline, offering a one-step quantification together with a principled classification of regulated features into distinct layers of gene regulation (Supplementary Methods S4 and Supplementary Table S1). In a first benchmark, we evaluated the detection performance of Ribotools, Riborex [26], anota2seq [27], and the deltaTE method [25] (Supplementary Methods S4 – benchmark 1). Ribotools performed on par with deltaTE and marginally outperformed



B

riboseq-results
- without-adapters
  - d1-1.fastq.gz
- with-rrna
- without-rrna
- without-rrna-mapping
  - d1-1.bam
- metagene-profiles
- orf-profiles
- orf-predictions
- count-tables
  - Ribo-d1-1.length-28-29-30.tsv
- analysis
  - profile_construction
  - rpbp_predictions

rnaseq-results
- without-adapters
- with-rrna
- without-rrna
- without-rrna-mapping
- count-tables
  - RNA-d1-1.length-30.tsv

tea-results
- LRT
  - d1_vs_d0
  - d5_vs_d1
- deltaTE
- sample-table.csv

Figure 1. Overview of the Ribotools box. (**A**) Ribotools, together with Rp-Bp, supports periodicity estimation and Ribo-seq QC, ORF discovery, abundance estimation (using either annotated genes or Ribo-seq ORFs as features), and TEA. Ribo-seq abundance is quantified using only periodic RPFs, and reads from matched RNA-seq libraries are trimmed to the maximum periodic fragment length to minimize mapping bias. The E2E pipeline supports the complete analysis workflow but can also be started from any intermediate step, with the option to skip completed steps or overwrite existing results. A single YAML configuration file is used throughout the workflow. These steps can be configured via the configuration file or using command-line arguments. Two interactive dashboards facilitate QC and Ribo-seq ORF exploration and can be deployed to share results with collaborators. Finally, Ribotools is built on a robust statistical framework and automatically classifies features into distinct layers of gene regulation. (**B**) Output files directory tree. The first directory is given in the configuration file, and the remaining tree is automatically generated and used to subsequently identify existing files on the system. Directories in red can be deleted upon completion to save on disk space, or kept for QC. Directories in light gray are only present when performing QC, when running the ORF discovery pipeline, or when using the deltaTE approach (LRT is the default). Required input and expected output for each step are described in the online documentation. This image has been designed using resources from Flaticon.com.

anota2seq in most test cases (Supplementary Results S1 and Supplementary Fig. S1). Riborex's apparent advantage in translation-specific comparisons reflected its undifferentiated TE-interaction score rather than genuine superiority; it performed worst when all regulatory classes were considered, owing to its inability to detect abundance-regulated genes. Classification accuracy against the ground-truth labels was similarly high and consistent across all classifying tools for the abundance and background classes, moderate for translation and buffering, with anota2seq showing comparatively lower recovery of translation genes (Supplementary Results S1 and Supplementary Fig. S2A and B). In a second benchmark, we used a subset of our newly generated dataset and performed label permutation to quantify Ribotools' empirical false positive rate (Supplementary Methods S4 – benchmark 2). This analysis yielded zero false positives at FDR < 0.05 under the null, supporting the specificity of the significant genes identified in the real data, despite the inherent limitation imposed by $n = 2$ replicates per condition (Supplementary Results S1 and Supplementary Fig. S2C). Finally, we performed a qualitative comparison of Ribotools with other available ORF discovery tools (Supplementary Results S2 and Supplementary Table S2). Among tools specifically designed for ORF calling and discovery, Ribotools/Rp-Bp is one of the few actively maintained options, with extensive online documentation and tutorials, comparable in this respect only to ORFik [28]. Ribotools and ORFik differ in functionality and implementation details, but only Ribotools/Rp-Bp provides interactive visualization dashboards. ORFik's ORF translation prediction relies on a heuristic, supervised learning approach, whereas Rp-Bp builds on a probabilistic model designed for unbiased *de novo* discovery, providing full posterior distributions rather than point classifications [17]. In summary, Ribotools emerged as the only E2E, modular, and actively maintained TEA tool, consistently achieving the best overall performance in our benchmark, and providing, in combination with Rp-Bp, a robust workflow for Ribo-seq data analysis, QC, and ORF discovery.

Transcriptome and translatome of hiPSC differentiation into cardiomyocytes

We used Ribotools to investigate how translational control evolves during cardiac lineage commitment *in vitro*. hiPSCs were differentiated into CMs and collected in duplicate at key developmental stages: hiPSCs (day 0), early mesoderm (day 1), cardiac mesoderm (day 5), cardiac progenitors (day 7), early CMs (day 17), and CMs (day 31). Each sample was subjected to transcriptome (RNA-seq) and translatome (Ribo-seq) profiling and analyzed with Ribotools.

Quality control

A demonstration of the QC dashboard for this dataset is hosted at https://rpbp.dieterichlab.org/hIPSCCMdiff/profile_construction. Each Ribo-seq library produced an average of 10.5 million mapped reads, of which ∼6.7 million periodic reads were retained for subsequent analyzes (Fig. 2A and online). The Ribo-seq data showed high quality; reads were predominantly 28 or 29 long, and an average of 84% were mapped to the primary reading frame of the annotated coding sequences (CDSs) (Fig. 2B and online). Length distributions and metagene profiles are available for each sample to inspect read length periodicity, enabling users to visualize characteristic signatures of translation such as post-initiation pausing at the fifth codon position [29] (Supplementary Fig. S3A and online).

ORF discovery

The Ribo-seq ORF discovery dashboard, available at https://rpbp.dieterichlab.org/hIPSCCMdiff/predictions, showcases alternative translation events that occur during cardiac lineage commitment (Fig. 2C and online). As expected, Ribo-seq ORFs were found mainly within Ensembl-annotated CDSs, but an average of 1861 upstream ORFs (uORFs), 1425 ncORFs, and ∼1218 variants of annotated CDSs were also translated at each developmental stage (Fig. 2D and online). Our *de novo* transcriptome allowed us to predict 392 completely novel, uncharacterized ORFs. A total of 1941 Ribo-seq ORFs were matched to the GENCODE consensus set of Ribo-seq ORFs (https://www.gencodegenes.org/pages/riboseq_orfs), of which ∼95% are unannotated small ORFs (sORFs), mostly uORFs, such as a GATA4 uORF recently described [1, 30, 31], or c12riboseqorf16, upstream of the glycolysis gene SLC2A3, identified by Ribotools as translationally downregulated after day 17 (Fig. 2E and F, Supplementary Fig. S3B and C). Additional results can be found in the Supplementary Results S3 and Supplementary Fig. S4.

Downstream analyses using ORF counts

Principal component analysis (PCA) from Ribo-seq ORF counts, obtained with the Ribotools abundance estimation pipeline, delineated the progression from hIPSCs to CMs in a comparable developmental order at both transcriptome and translatome levels (Fig. 2G). Cells on day 0 and day 1 showed highly similar profiles, marked by the expression of pluripotency and early mesoderm markers (Fig. 2H). At days 5 and 7, cells progressed from cardiac-committed mesoderm to cardiac progenitors. From day 17 onward, they predominantly expressed CM markers. In general, the abundance profiles of cardiac-committed hiPSCs revealed a transition from low MYH7:MYH6, TNII3:TNNI1, MYL2:MYL7, and IRX4:NR2F2 on days 5 and 7, to higher ratios on days 17 and 31, confirming a shift from fetal to mature-like CMs (Fig. 2I).

In summary, we obtained a global snapshot of RNA translation across hiPSC differentiation into CMs, and the Ribotools box allowed us to demonstrate that hiPSC-CMs are a reasonable model to better understand cardiac lineage commitment.

Translational control during hiPSC differentiation into cardiomyocytes

Using Ribotools, we were able to show how translational regulation follows a dynamic trajectory during CM differentiation, with multiple layers for commitment, execution, control, and stability (Fig. 3 and Supplementary Fig. S5). On day 1, coordinated changes mediated a wide spectrum of cellular responses crucial for cell differentiation, cell fate, and development in general (Fig. 3A). On day 5, cells experienced a major rewiring as they exited pluripotency and underwent lineage commitment. During early CM differentiation, genes involved in translation were predominantly translationally downregulated, consistent with coordinated repression of the translation machinery during the shift from proliferative to lineage-specific programs (Fig. 3B). These observations were supported by gene set enrichment analysis (GSEA), where we observed a negative enrichment for rRNA processing (NES

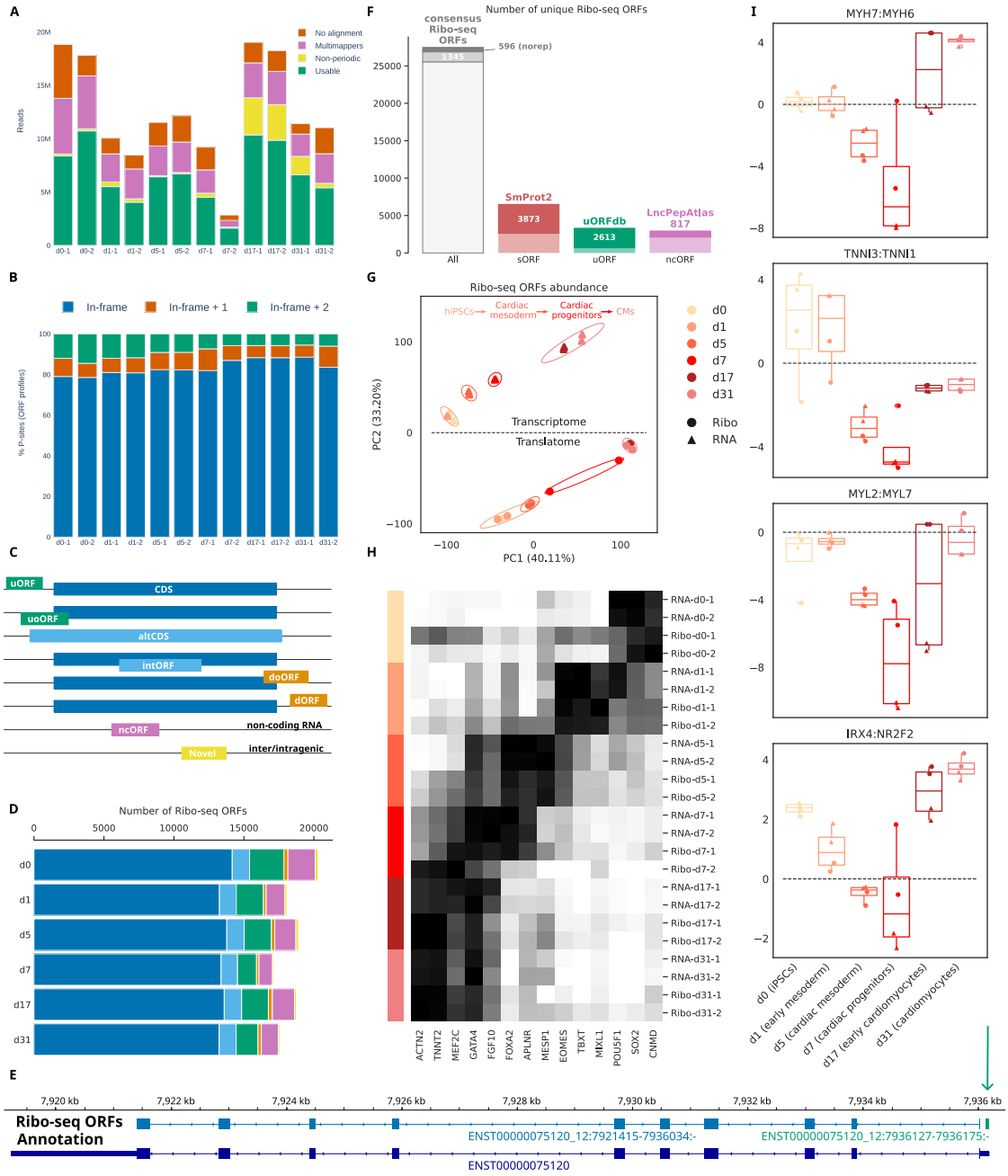


Figure 2. Transcriptome and translatome profiling capture cardiac lineage commitment and identifies unannotated Ribo-seq ORFs during hiPSC-differentiation. (**A**) Image taken from the QC dashboard. Stacked bars showing the number of reads classified as usable, non-periodic, multimapping, or not aligned, for each Ribo-seq library. Read filtering can also show the number of reads after *in-silico* rRNA removal for experimental protocol assessment. (**B**) Image taken from the QC dashboard showing the proportion of periodic RPFs assigned to each reading frame. (**C**) Image adapted from the ORF discovery dashboard. Schematic representation of the Ribo-seq ORF categories assigned according to their positions relative to annotated CDSs: upstream ORF (uORF), upstream overlapping ORF (uoORF), alternative CDS (altCDS), internal ORF (intORF), downstream overlapping ORF (doORF), downstream ORF (dORF), ORF in a transcript annotated as noncoding (ncORF), and novel intergenic or intragenic ORF (novel). (**D**) Bars with the total number of Ribo-seq ORFs predicted by Rp-Bp during hiPSC development to CMs, divided by developmental stage and ORF category. (**E**) Image adapted from the ORF discovery dashboard showing the integrated IGV browser and a predicted uORF upstream of the annotated SLC2A3 CDS. Rp-Bp assigned this ORF the identifier ENST00000075120 12:7936127–7936175:-. The ORFs are optionally assigned an extra identifier (c12riboseqorf16), matching that of the GENCODE consensus set of Ribo-seq ORFs, searchable via the ORF discovery dashboard. (**F**) Total number of Ribo-seq ORFs found during hiPSC development to CMs with evidence from the GENCODE consensus set of Ribo-seq ORFs (including non-replicated, or "norep" ORFs, found in only one experiment, i.e. not confirmed by any other independent research at the time), SmProt2 (for sORFs), uORFdb (for uORFs), and LncPepAtlas (for ncORFs). The numbers of top of each bar show how many Ribo-seq ORFs were found in each database, per category. Standardized ORFs were matched by coordinates; SmProt2, uORFdb, and LncPepAtlas were matched by amino acid sequence. (**G**) PCA of the RNA- and Ribo-seq abundances obtained with Ribotools for all Ribo-seq ORFs found in this study. Differentiation stages are colored from day 0 to day 31. Triangles represent RNA-seq and circles represent Ribo-seq samples. (**H**) Heatmap with scaled abundances for selected pluripotency, mesodermal, cardiac progenitor, and cardiac marker genes using variance-stabilized gene counts obtained with Ribotools. Differentiation stages are colored as in panel (G). (**I**) Box plots with MYH7:MYH6, TNNI3:TNNI1, MYL2:MYL7, and IRX4:NR2F2 ratios using variance-stabilized gene counts obtained with Ribotools across developmental stages (same legend as in panel G). Boxes indicate interquartile range; center line indicates median.

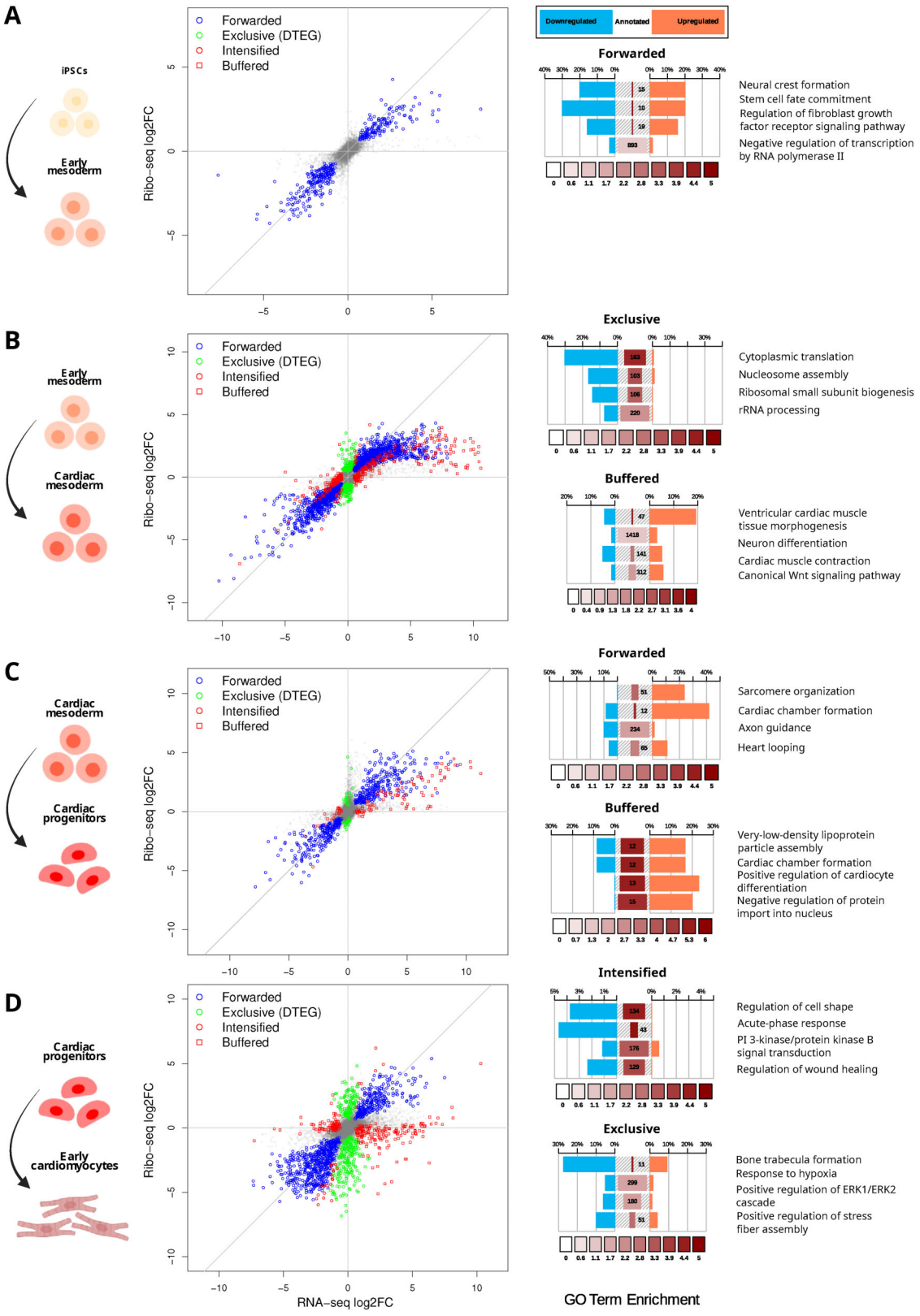


Figure 3. Transcription and translation are highly dynamic during differentiation. From top to bottom, the comparisons show (**A**) day 1 versus day 0 (from hiPSCs to early mesoderm), (**B**) day 5 versus day 1 (from early to cardiac mesoderm), (**C**) day 7 versus day 5 (from cardiac mesoderm to progenitor cells), and (**D**) day 17 versus day 7 (from cardiac progenitors to early CMs). Left: Schematic representation of the corresponding differentiation transition. Middle: Output from Ribotools. Scatter plots comparing RNA-seq $\log_2$FC (*x*-axis) to Ribo-seq log2FC (*y*-axis) for each gene. Genes were classified by Ribotools as forwarded, differential-translation exclusive (DTEG), intensified, or buffered. Right: Gene Ontology (Biological Process) over-representation analyses for selected classes, showing the proportion of differentially down- and up-regulated genes. The central column and associated color bar show the Gene Ontology term log enrichment score, together with the number of annotated genes. All details are in the Supplementary Methods S5. See also Supplementary Fig. S5.

−2.8), translation elongation (NES −2.6), translation initiation (NES −2.6), and cytoplasmic translation (NES -2.2), in which an important number of translationally downregulated ribosomal protein genes were part of the core enrichment set (Supplementary Methods S5 on functional enrichment analyses and Supplementary Table S3). These changes may reflect a shift in translational capacity and selectivity as cell acquire lineage-specific translational programs. Whether this transition involves altered ribosome composition or specialization remains to be determined [32, 33]. At the same time, key developmental cardiac pathways were transcriptionally activated, but translationally buffered (Fig. 3B and Supplementary Fig. S5A), suggesting a "timing and dosage" mechanism for fine-tuning cardiac protein expression, enabling a more precise temporal control than would be achievable through transcription alone. Compared to day 0, buffered pathways on day 5, as highlighted by GSEA, were enriched for the morphogenesis of cardiac septum (NES 2.3), cardiac muscle tissue (NES 2.2), and endocardial cushion (NES 2.1). These observations were partly confirmed by differential proteomics (Supplementary Methods S6 on protein correlation analysis and Supplementary Figs S6 and S7A–C), where upregulated proteins were still largely enriched for alternative lineage programs. By day 7, cardiac and myogenic genes transitioned from buffered to forwarded regulation, indicating lineage restriction and active execution of the differentiation program at the translational (Fig. 3C) and protein level (Supplementary Fig. S6 and Supplementary Fig. S7D and E). Persistent buffering of cardiac, lipid metabolic, and nuclear transport processes suggested ongoing fine-tuning and staged activation of functional and metabolic networks. As cells transitioned to maturation on day 17, they moved from a responsive/adaptive state to a more stable functional state, actively reducing remodeling and signaling plasticity, and shutting down alternative fate decision programs at the translational level (Fig. 3D). Although we observed refinement and reorganization at the translational level, changes in protein abundance remained predominantly associated with cardiac lineage and functional programs (Supplementary Fig. S6 and Supplementary Fig. S7E). On day 31, the system had reached a stable state in which protein output appeared to be actively maintained rather than dynamically remodeled (Supplementary Figs S5B, S6, and S7F). At this stage, and in the absence of changing conditions, translational regulation primarily serves to maintain protein homeostasis, with most pathways remaining stable through buffering mechanisms. Specifically, changes in RNA abundance were compensated by translation efficiency (i.e. a significant interaction term), resulting in stable RPF levels despite significant transcript-level changes (Supplementary Fig. S5B). Across most developmental stages, the differential proteomics were broadly concordant with the expectations of the Ribotools classification (Supplementary Methods S5, Supplementary Results S4, and Supplementary Fig. S8).

Discussion

The combined analysis of RNA- and Ribo-seq data is challenging and requires reliable tools. Here, we present the Ribotools box, an open-source, configuration-based, and modular package underpinned by Rp-Bp [17]. In contrast to naive ratio-based approaches [34, 35], changes in TE are inferred by testing the interaction term in a negative-binomial GLM fitted to RNA- and Ribo-seq counts [25, 26, 28]. Building on this established framework ensures methodological reliability while addressing practical gaps in the TEA landscape: dedicated QC tools focus on QC and visualization rather than complete analyses [36, 37], whereas TEA-only tools are designed for ready-to-use count data and lack Ribo-seq-specific QC; moreover, most are now deprecated or obsolete (refer to Supplementary Methods S4—Qualifying criteria—and Supplementary Table S1 for references). Ribotools is among the few well-documented and actively maintained E2E tools, providing QC, ORF discovery, and TEA "out-of-the-box" through an easy-to-use yet robust statistical framework. Like many widely adopted bioinformatics pipelines, its value lies not in introducing entirely new algorithms, but in providing a coherent and practical E2E solution.

Future work could include improved QC metrics and visualization, an interactive TEA dashboard, and support for specific protocols. Despite these limitations, Ribotools emerged as the strongest all-around performer, both in detection and classification and in the broader comparison of features across available TEA and ORF discovery tools (refer to Supplementary Results S2—Qualifying criteria and Ribotools stands out as the best all-around performer—and Supplementary Table S2 for additional references). Given the extensive evaluations of TEA methods in previous studies [25–27], our benchmark was not intended as another exhaustive comparison, but as a streamlined, pragmatic, and evolvable framework. The complete workflow is openly available and fully reproducible.

The active role of translation in cell-fate decisions during human cardiogenesis has long been recognized. More recently, Ribo-seq has been increasingly used to investigate cardiac reprogramming and to characterize the cardiac translatome at scale [1, 7, 8, 33]. Here, we demonstrate how Rp-Bp and Ribotools streamline the analysis of complex datasets and provide insights into the mechanisms underlying cardiac lineage commitment. We acknowledge that proteomics support for Ribo-seq ORFs remains limited, a well-recognized challenge, in particular for sORFs [14, 15]. Our objective is to demonstrate how Ribotools not only simplifies translatome profiling but also facilitates the early stages of ORF discovery. These extensive data represent a powerful resource to interrogate hiPSC-CMs as a model system, providing a source of potentially novel regulatory uORFs, alternative translation events, and much more.

Acknowledgements

The authors acknowledge the development work on Rp-Bp by the Scientific Software Center of Heidelberg University, in particular, by Liam Keegan and Dominic Kempf. The graphical abstract was created in BioRender. Dieterich, C. (2026) https://BioRender.com/klxzmja.

Author contributions: Etienne Boileau (Conceptualization, Software, Methodology, Investigation, Writing—original draft), Maja Bencun (Conceptualization, Experimentation, Investigation, Writing—review and editing), Francesca Tuorto (Experimentation, Writing—review and editing), Isabel S. Naarmann-de Vries (Experimentation, Writing—review and editing), Pauline Fahjen (Experimentation, Writing—review), Philipp Mertins (Resources, Writing—review), and Christoph Dieterich (Funding acquisition, Resources, Project administra-

tion, Conceptualization, Writing—review and editing). All authors read and approved the final manuscript.

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

This project was funded by the Klaus Tschira Stiftung gGmbH (grant #00.013.2021).

Data availability

Raw sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under accession number PRJNA1447392. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD076601. Additional data supporting this study are available on Zenodo at 10.5281/zenodo.17899627. The Rp-Bp source code is available at https://github.com/dieterich-lab/rp-bp or https://zenodo.org/records/17974359. The Ribotools source code is available at https://github.com/eboileau/ribotools or https://zenodo.org/records/17974316. The source for the benchmarking pipeline is available at https://github.com/eboileau/ribotools-benchmark or https://zenodo.org/records/21409454. Packages and containers are available on Bioconda at https://bioconda.github.io/recipes/rpbp/README.html and https://bioconda.github.io/recipes/ribotools/README.html.

References

1. van Heesch S, Witte F, Schneider-Lunitz V *et al.* The translational landscape of the human heart. *Cell* 2019;178:242–60.e29. https://doi.org/10.1016/j.cell.2019.05.010
2. Jordan B, Weidenbach K, Schmitz RA. The power of the small: the underestimated role of small proteins in bacterial and archaeal physiology. *Curr Opin Microbiol* 2023;76:102384. https://doi.org/10.1016/j.mib.2023.102384
3. Weingarten-Gabbay S, Bauer MR, Stanton AC *et al.* Pan-viral ORFs discovery using massively parallel ribosome profiling. *Science* 2025;388:1218–24. https://doi.org/10.1126/science.ado6670
4. Chothani SP, Adami E, Widjaja AA *et al.* A high-resolution map of human RNA translation. *Mol Cell* 2022;82:2885–99. https://doi.org/10.1016/j.molcel.2022.06.023
5. Sandmann C-L, Schulz JF, Ruiz-Orera J *et al.* Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. *Mol Cell* 2023;83:994–1011. https://doi.org/10.1016/j.molcel.2023.01.023
6. Schafer S, Adami E, Heinig M *et al.* Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nat Commun* 2015;6:7200. https://doi.org/10.1038/ncomms8200
7. Xie Y, Wang Q, Yang Y *et al.* Translational landscape of direct cardiac reprogramming reveals a role of Ybx1 in repressing cardiac fate acquisition. *Nat Cardiovasc Res* 2023;2:1060–77. https://doi.org/10.1038/s44161-023-00344-5
8. Ruiz-Orera J, Miller DC, Greiner J *et al.* Evolution of translational control and the emergence of genes and open reading frames in human and non-human primate hearts. *Nat Cardiovasc Res* 2024;3:1217–35. https://doi.org/10.1038/s44161-024-00544-7
9. Olexiouk V, Van Criekinge W, Menschaert G. An update on sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res* 2017;46:D497–502. https://doi.org/10.1093/nar/gkx1130
10. Hao Y, Zhang L, Niu Y *et al.* SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief Bioinform* 2017;19:636–43.
11. Manske F, Ogoniak L, Jürgens L *et al.* The new uORFdb: integrating literature, sequence, and variation data in a central hub for uORF research. *Nucleic Acids Res* 2022;51:D328–36. https://doi.org/10.1093/nar/gkac899
12. Tierney JAS, Świrski MI, Tjeldnes H *et al.* RiboSeq.Org: an integrated suite of resources for ribosome profiling data analysis and visualization. *Nucleic Acids Res* 2024;53:D268–74. https://doi.org/10.1093/nar/gkae1020
13. Zhou X, Qin Y, Li J *et al.* LncPepAtlas: a comprehensive resource for exploring the translational landscape of long non-coding RNAs. *Nucleic Acids Res* 2024;53:D468–76. https://doi.org/10.1093/nar/gkae905
14. Mudge JM, Ruiz-Orera J, Prensner JR *et al.* Standardized annotation of translated open reading frames. *Nat Biotechnol* 2022;40:994–99. https://doi.org/10.1038/s41587-022-01369-0
15. Deutsch EW, Kok LW, Mudge JM *et al.* Expanding the human proteome with microproteins and peptideins. *Nature* 2026;654:813–25. https://doi.org/10.1038/s41586-026-10459-x
16. Wang H, Wang Y, Xie Z. Computational resources for ribosome profiling: from database to Web server and software. *Brief Bioinform* 2017;20:144–55. https://doi.org/10.1093/bib/bbx093
17. Malone B, Atanassov I, Aeschimann F *et al.* Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res* 2017;45:2960–72.
18. Kmietczyk V, Riechert E, Kalinski L *et al.* m6A-mRNA methylation regulates cardiac gene expression and cellular growth. *Life Sci Alliance* 2019;2:e201800233. https://doi.org/10.26508/lsa.201800233
19. Górska AA, Sandmann C, Riechert E *et al.* Muscle-specific Cand2 is translationally upregulated by mTORC1 and promotes adverse cardiac remodeling. *EMBO Rep* 2021;22:e52170.
20. Froese N, Cordero J, Abouissa A *et al.* Analysis of myocardial cellular gene expression during pressure overload reveals matrix based functional intercellular communication. *iScience* 2022;25:103965. https://doi.org/10.1016/j.isci.2022.103965
21. Ouspenskaia T, Law T, Clauser KR *et al.* Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat Biotechnol* 2022;40:209–17. https://doi.org/10.1038/s41587-021-01021-3
22. Trendel J, Boileau E, Jochem M *et al.* PEPseq quantifies transcriptome-wide changes in protein occupancy and reveals selective translational repression after translational stress. *Nucleic Acids Res* 2023;51:e79. https://doi.org/10.1093/nar/gkad557
23. Kleinsorge M, Cyganek L. Subtype-directed differentiation of human iPSCs into atrial and ventricular cardiomyocytes. *STAR Protoc* 2020;1:100026. https://doi.org/10.1016/j.xpro.2020.100026
24. Tuorto F, Herbst F, Alerasool N *et al.* The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *EMBO J* 2015;34:2350–62. https://doi.org/10.15252/embj.201591382
25. Chothani S, Adami E, Ouyang JF *et al.* deltaTE: detection of translationally regulated genes by integrative analysis of Ribo-seq and RNA-seq data. *Curr Protoc Mol Biol* 2019; 129: e108. https://doi.org/10.1002/cpmb.108
26. Li W, Wang W, Uren PJ *et al.* Riborex: fast and flexible identification of differential translation from Ribo-seq data.

Bioinformatics 2017;33:1735–37. https://doi.org/10.1093/bioinformatics/btx047
27. Oertlin C, Lorent J, Murie C *et al.* Generally applicable transcriptome-wide analysis of translation using anota2seq. *Nucleic Acids Res.* 2019;47:e70. https://doi.org/10.1093/nar/gkz223
28. Tjeldnes H, Labun K, Torres Cleuren Y *et al.* ORFik: a comprehensive R toolkit for the analysis of translation. *BMC Bioinformatics* 2021;22:336. https://doi.org/10.1186/s12859-021-04254-w
29. Han Y, Gao X, Liu B *et al.* Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. *Cell Res.* 2014;24:842–51. https://doi.org/10.1038/cr.2014.74
30. Doroudgar S, Hofmann C, Boileau E *et al.* Monitoring cell-type-specific gene expression using ribosome profiling in vivo during cardiac hemodynamic stress. *Circ Res* 2019;125:431–48. https://doi.org/10.1161/CIRCRESAHA.119.314817
31. Hedaya OM, Jiang F, Baliga U *et al.* Upstream open reading frame inactivation augments GATA4 translation and cardiomyocyte hypertrophy in mice. bioRxiv, 2025.
32. Shiraishi C, Matsumoto A, Ichihara K *et al.* RPL3L-containing ribosomes determine translation elongation dynamics required for cardiac function. *Nat Commun* 2023;14:2131.
33. Bartsch D, Kalamkar K, Ahuja G *et al.* mRNA translational specialization by RBPMS presets the competence for cardiac commitment in hESCs. *Sci Adv* 2023;9:eade1792.
34. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 2011;147:789–802.
35. Thoreen CC, Chantranupong L, Keys HR *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* 2012;485:109–13.
36. Lauria F, Tebaldi T, Bernabò P *et al.* riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput Biol* 2018;14:1–20.
37. Legrand C, Tuorto F. RiboVIEW: a computational framework for visualization, quality control and statistical analysis of ribosome profiling data. *Nucleic Acids Res* 2020;48:e7.

Received: April 9, 2026. **Revised:** September 11, 2026. **Accepted:** September 15, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.