Nucleic Acids Research, 2025, **53**, D293–D298
https://doi.org/10.1093/nar/gkae808
Advance access publication date: 25 September 2024
Database issue

RPFdb v3.0: an enhanced repository for ribosome profiling data and related content

Yan Wang†, Yuewen Tang†, Zhi Xie* and Hongwei Wang*

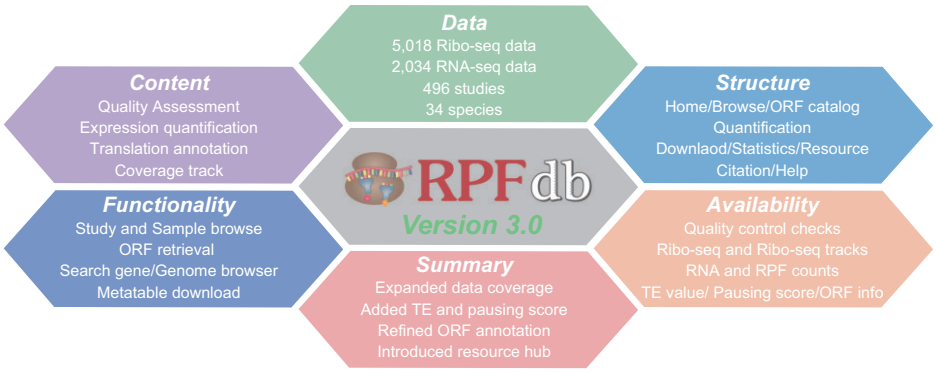
State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangzhou 510060, China

*To whom correspondence should be addressed. Tel: +86 20 6667 7086; Email: xiezhi@gmail.com
Correspondence may also be addressed to Hongwei Wang. Tel: +86 20 6667 7086; Email: bioccwhw@126.com
†The first two authors should be regarded as Joint First Authors.

Abstract

RPFdb (http://www.rpfdb.org or http://sysbio.gzzoc.com/rpfdb/) is a comprehensive repository dedicated to hosting ribosome profiling (Ribo-seq) data and related content. Herein, we present RPFdb v3.0, a significant update featuring expanded data content and improved functionality. Key enhancements include (i) increased data coverage, now encompassing 5018 Ribo-seq datasets and 2343 matched RNA-seq datasets from 496 studies across 34 species; (ii) implementation of translation efficiency, combining Ribo-seq and RNA-seq data to provide gene-specific translation efficiency; (iii) addition of pausing score, facilitating the identification of condition-specific triplet amino acid motifs with enhanced ribosome enrichment; (iv) refinement of open reading frame (ORF) annotation, leveraging RibORF v2.0 for more sensitive detection of actively translated ORFs; (v) introduction of a resource hub, curating advances in translatome sequencing techniques and data analytics tools to support a panoramic overview of the field; and (vi) redesigned web interface, providing intuitive navigation with dedicated pages for streamlined data retrieval, comparison and visualization. These enhancements make RPFdb a more powerful and user-friendly resource for researchers in the field of translatomics. The database is freely accessible and regularly updated to ensure its continued relevance to the scientific community.

Graphical abstract



Introduction

Translation is a highly regulated process that plays a crucial role in shaping cellular proteomes (1). Its significance extends beyond protein synthesis, serving as a rapid and reversible means of gene regulation that enables cells to respond dynamically to environmental cues and stress conditions (2). The multi-step nature of translation, including initiation, elongation and termination phases, each regulated by numerous factors, underscores its complexity (3). Recent years have seen a paradigm shift in our understanding of gene translation and translational regulation. A pivotal advancement in this field is the development of ribosome profiling (Ribo-seq) by Ingolia and Weissman in 2009 (4). This breakthrough technique revolutionized translatome research, providing an unprecedented, genome-wide, high-resolution view of messenger RNA (mRNA) translation by capturing and sequencing ribosome-protected mRNA fragments.

Since its inception, Ribo-seq has undergone significant methodological refinements to address technical challenges such as ribosome runoff during sample preparation and biases in library construction (5). These improvements, coupled with advances in bioinformatics (6), have broadened its applications. Consequently, Ribo-seq has facilitated groundbreaking discoveries: (i) identification of pervasive translation

Received: July 22, 2024. **Revised:** August 15, 2024. **Editorial Decision:** September 2, 2024. **Accepted:** September 5, 2024

outside annotated protein-coding regions, including within long non-coding RNAs and circular RNAs, challenging traditional definitions of the coding genome (7–9); (ii) revelation of the prevalence and importance of upstream and downstream open reading frames (ORFs) and alternative start codons in translational control, elucidating novel mechanisms of gene expression regulation (10,11); (iii) insights into phenomena such as ribosome pausing and its role in protein folding and regulation, linking translational kinetics to protein function (12); (iv) elucidation of translational dysregulation in various pathologies, opening new avenues for therapeutic interventions (13); and (v) understanding of how viruses hijack and manipulate host translation machinery, informing the development of novel antiviral strategies (14).

These discoveries have precipitated a surge in Ribo-seq experiments across molecular biology and biomedicine. As researchers increasingly embrace this powerful technique, the growing volume of data has necessitated the creation of dedicated databases for hosting, processing and analyzing this information. These databases range from comprehensive, multi-species repositories such as RPFdb (15), Trips-Viz (16), GWIPS-viz (17) and TranslatomeDB (18) to more specialized resources such as Ribo-uORF (19), uORFdb (20), riboCIRC (21) and sORFs.org (22). The development of these databases has significantly reduced technical barriers, democratizing access to this complex data type and enabling researchers without extensive bioinformatics expertise to explore and utilize it effectively.

RPFdb is purpose-built to host Ribo-seq data and related content, processed through a unified pipeline to ensure consistency and comparability across datasets (15,23). To maintain its relevance and utility in this rapidly advancing field, RPFdb now undergoes systematic updates. This iterative process involves not only the addition of new datasets but also the refinement of ORF annotation, incorporation of new analytical contents and enhancement of user interfaces based on user feedback and emerging requirements. By continuously improving and expanding RPFdb, our objective is to provide the ribosome profiling community with a comprehensive, up-to-date resource, facilitating translation of raw sequencing data into meaningful biological insights.

Summary of RPFdb features and functionalities

Since its initial release in 2016 (23), RPFdb has evolved to version 3.0, now offering enhanced capabilities for the exploration of Ribo-seq data and related content. Figure 1 illustrates the key features of RPFdb v3.0, showcasing its comprehensive functionalities for researchers. RPFdb v3.0 maintains its intuitive interface while significantly expanding its functionalities: (i) Browse: enables users to explore an extensive collection of studies and samples with detailed information on data descriptions and data quality assessment; (ii) ORF catalog: presents an updated compilation of actively translated ORFs with enhanced annotation and filtering options; (iii) Quantification: facilitates the visualization of footprints and comparative analysis of RPKM (Reads Per Kilobase per Million mapped reads) values across genomic regions, as well as the assessment of translation efficiency and pausing scores; (iv) Download: provides access to tabular metadata and the latest Ribo-seq and RNA-seq tracks; (v) Statistics: offers graphical summaries of collected data; (vi) Resource: informs users about cutting-edge advances in translatome sequencing techniques and data analytical tools; (vii) Citation: guides proper attribution and usage of the database and (viii) Help: includes a step-by-step tutorial for new users. These enhanced features collectively empower RPFdb to promote collaboration and knowledge sharing within the ribosome profiling community.

Expansion and improvement of the RPFdb database

Increased data coverage

The evolution of RPFdb into its current version, RPFdb v3.0, represents a significant advancement, particularly in terms of data coverage and diversity (Table 1). The database now hosts an extensive collection of 5018 Ribo-seq datasets derived from 496 distinct studies. These datasets encompass a wide spectrum of experimental conditions, tissue types, developmental stages and environmental stimuli, providing researchers with a rich array of biological contexts to explore. A key enhancement in RPFdb v3.0 is the inclusion of 2343 matched RNA-seq datasets. This integration allows for comparative analyses that link translational dynamics with transcriptomic profiles, offering a more comprehensive view of gene expression regulation. The database's scope has expanded to cover 34 different species, from unicellular microbes to multicellular plants and animals, underscoring its broad applicability across diverse biological systems.

The progression from RPFdb v1.0 to v3.0 demonstrates consistent and substantial growth. Each version has markedly increased the number of samples, studies and species represented. Notably, the latest version nearly doubles the Ribo-seq datasets compared with v2.0 and represents more than a 6-fold increase from v1.0. The introduction of matched RNA-seq data further enhances the database's utility for comprehensive translational studies. This expansion not only increases the quantity of data available but also improves the quality and depth of potential analyses. The broader species coverage facilitates comparative studies across evolutionarily diverse organisms, while the inclusion of matched RNA-seq data enables researchers to distinguish between transcriptional and translational regulation effects.

Enhanced ORF annotation

Actively translated ORFs represent the regions of the genome that are transcribed into mRNA and subsequently translated into proteins. Accurate annotation of actively translated ORFs is crucial for understanding the fundamental mechanisms of gene expression and protein synthesis. We refine actively translated ORF annotations through the use of RibORF v2.0 software (24,25), which automates data quality control, selects 3-nt periodic reads and employs ribosomal A-site corrected reads to identify genome-wide ORFs. To improve the reliability of ORF identification among duplicate samples, we have replaced the previous method of combining Ribo-seq data with duplicate samples. This change allows for a more accurate assessment of ORF consistency across duplicates, providing researchers with more reliable data. In addition to these refined annotations, we offer a consensus set of Ribo-seq ORFs for each species. This consensus set represents a standardized collection by merging RibORF-identified ORFs from different conditions (26). By providing refined ORF annota-

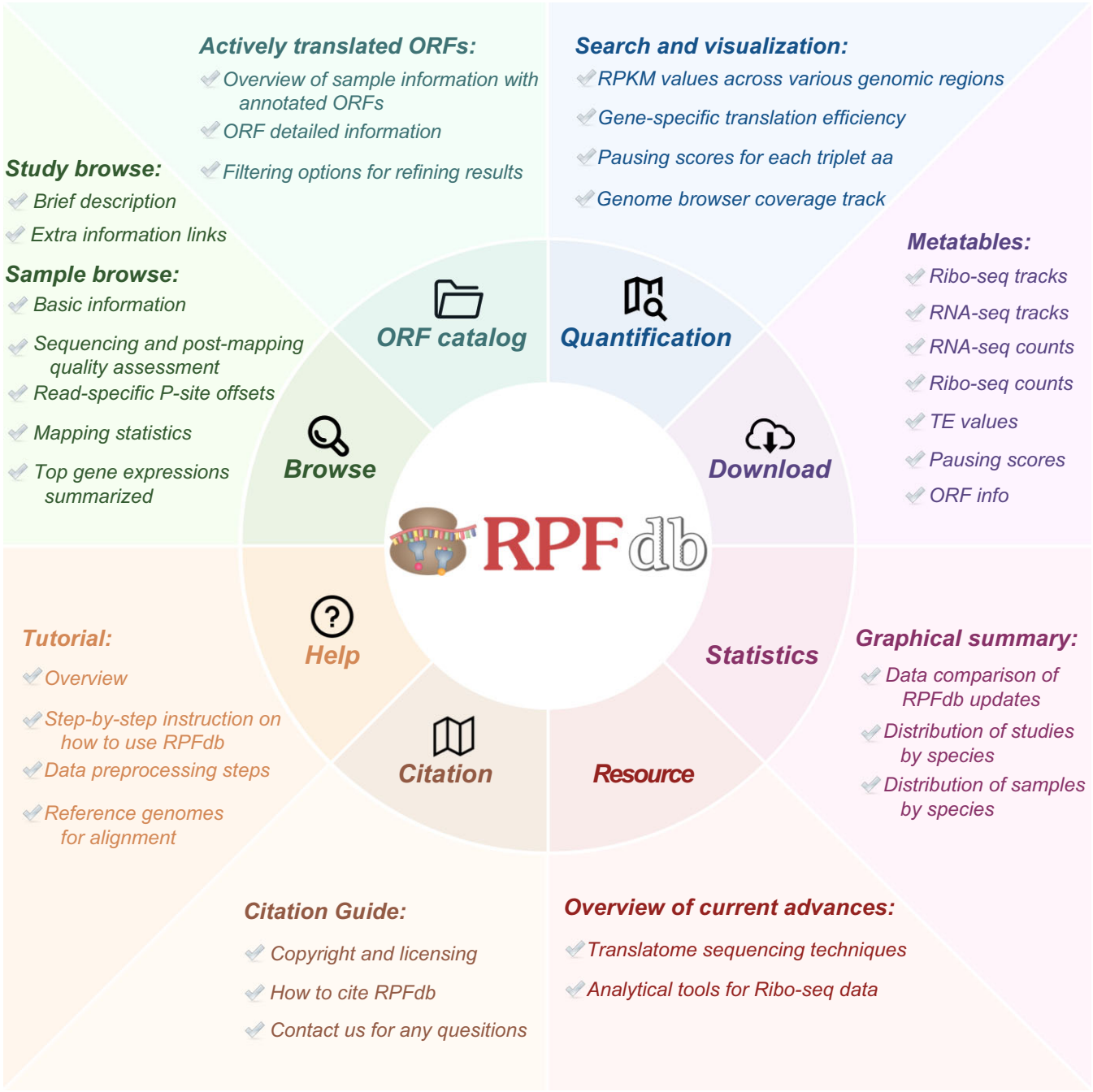


Figure 1. Overview of RPFdb features and functionalities.

tions and consensus sets, RPFdb v3.0 not only enhances the accuracy of translational data but also supports more consistent and reproducible research outcomes across different studies and species.

New content presentation

Optimizing primary sequences to enhance mRNA translation is an important focus in the development of mRNA-based therapeutics (27). Translation efficiency and ribosome pausing are known to be two critical factors influencing the dynamics of translation. In this update, we tackle these aspects in RPFdb v3.0, facilitating deeper insights into the intricacies of mRNA translation. We now present translation efficiency estimates by dividing the footprint RPKM by mRNA RPKM for each gene's coding sequence, utilizing both Ribo-seq and RNA-seq data (4). This enables researchers to quantitatively assess the rate of mRNA translation into proteins, identifying genes with high translation activity and shedding light on the cellular mechanisms governing protein production. Additionally, we introduce context-specific pausing scores for each triplet amino acid (tri-AA). These scores, calculated as the sum of normalized ribosome densities on each tri-AA motif using RiboMiner software (28), are crucial for pinpointing tri-AA motifs with enhanced ribosome enrichment. Such motifs often signify regions where translation is temporarily halted or

Table 1. Summary of RPFdb updates

	Version 1.0	Version 2.0	Version 3.0
Release date	2016	2019	2024
Data			
Type	Ribo-seq only	Ribo-seq only	Ribo-seq + RNA-seq
Species	8	29	34
Studies	82	293	496
Samples	777	2884	5018 + 2343
Content			
Quality assessment	• Sequencing QC report	• Sequencing QC report	• Sequencing QC report • Post-mapping QC report
Expression quantification	• Gene translation (RPKM)	• Gene translation (RPKM, RawCount)	• Gene translation (RPKM, RawCount) • Gene transcription (RawCount) • Translation efficiency • Pausing score
Translation annotation		• ORF catalog	• ORF catalog • Consensus Ribo-seq ORF set
Coverage track			• Ribo-seq tracks (bigWig) • RNA-seq tracks (bigWig)

slowed, potentially influencing protein folding, localization or regulatory functions. These new features will improve our understanding of translational dynamics by elucidating where and how ribosomes regulate protein synthesis. Furthermore, Ribo-seq tracks coupled with RNA-seq tracks are currently available, enabling intuitive display of coverage signals over genomic ranges and supporting comparative analysis of translation rates. Potentially, this update has significant implications for developing RNA therapeutic strategies, such as aiding in the design of mRNA sequences with optimized translation rates and providing a basis for identifying and improving translational pause sites.

New resource webpage

We have added a resource webpage to RPFdb v3.0 that serves as an information hub for researchers engaged in translatome studies, offering a concise yet informative summary of recent advancements in this rapidly evolving field. This curated collection highlights cutting-edge techniques that have revolutionized our ability to investigate translation at a genome-wide scale. Central to this resource is an overview of Ribo-seq and its variant methodologies, ranging from bulk profiling to single-cell profiling and spatial profiling. The page also features a section on computational tools specifically designed for Ribo-seq data analysis, covering software for read alignment, differential translation analysis and visualization of ribosome occupancy profiles, among others. By consolidating this information, we aim to equip researchers with essential knowledge to advance their translational research endeavors. This centralized resource not only facilitates easier access to cutting-edge methodologies but also promotes best practices in the field of translatome research.

Improved web interface

The significant expansion in both datasets and content necessitated an enhanced web interface. To achieve easier navigation, we redesigned the study browse and sample browse pages, offering a more intuitive layout and improved user experience. Furthermore, we optimized the ORF search response on the ORF catalog page, significantly reducing wait times for users. Additionally, we introduced new dedicated pages for quantification content, including measurements of ribosome occupancy, translation efficiency and ribosome pausing. These separate pages streamline the processes of data retrieval, comparison and visualization, thereby enhancing researchers' capability to derive meaningful insights efficiently. To facilitate secondary analyses of data, we also enhanced the download functionality. The download page now supports retrieval of translation efficiency, pausing score, Ribo-seq tracks and RNA-seq tracks. These improvements collectively aim to provide a more user-friendly and comprehensive platform for researchers engaged in translatome studies, supporting more efficient data access and analysis.

Database usage example

To access information about specific Ribo-seq data and its related content, users can start by navigating to the 'Study browse' page under the 'Browse' button on the homepage. For instance, entering the keyword 'eye' in the search box will return a dataset titled 'Change in translation efficiency in mouse eyes at P0.5 by RNG140 knockout'. Clicking the 'Details' button will direct users to the 'Sample browse' page, which provides comprehensive meta information about the dataset, including sample attributes and experimental variables, quality control checks for raw sequence data and post-mapping data, mapping statistics for each sample and more. For this dataset, users can review actively translated ORFs on the 'ORF catalog' page, and quantitative measurements on the 'Ribosome occupancy', 'Translation efficiency' and 'Ribosome pausing' pages under the 'Quantification' button. Notably, the 'Ribosome occupancy' and 'Translation efficiency' pages allow users to retrieve individual gene information. For example, by selecting 'M.musculus' as the species and entering 'Smad4 (a gene important in eye development and disease) as the gene of interest, users can compare results across different experimental conditions within the same species. The 'Ribosome occupancy' page also features a genome browser for visualizing Ribo-seq tracks, enabling users to examine the distribution of ribosomes along transcripts. All the quantitative measurements for this dataset can be downloaded from the 'Download' page, facilitating further analysis and integration with other data types. This example demonstrates how

users can efficiently navigate the database to access, visualize and analyze Ribo-seq data relevant to their research interests.

Conclusion and discussion

The RPFdb database is dedicated to advancing research within the ribosome profiling community, providing a comprehensive platform to explore Ribo-seq data and related content across diverse species. The latest update to RPFdb has further enhanced its utility with expanded data content and improved functionality, facilitating more efficient and insightful analyses. Although RPFdb has made substantial strides in democratizing access to Ribo-seq data and related contents, continuous development and adaptation will be crucial to keep pace with the rapid evolution in the field of translatomics. As Ribo-seq technology continues to advance, particularly with the emergence of single-cell translatomics (29–31), RPFdb must adapt to accommodate these new data types. Including single-cell Ribo-seq data would allow researchers to explore translational heterogeneity, providing unprecedented insights into cell-specific translation and translational regulation. Moreover, integrating spatial Ribo-seq data could offer a new dimension to our understanding of localized translation within tissues. Additional development of RPFdb could focus on incorporating more sophisticated online analysis and visualization tools directly into the platform. This could include interactive data exploration features, allowing users to perform custom secondary analyses without the need for local computational resources. Advanced visualization tools could enable researchers to generate publication-quality figures directly from the web interface, enhancing the accessibility and interpretability of complex Ribo-seq data.

Data availability

RPFdb is publicly available at http://www.rpfdb.org or http://sysbio.sysu.edu.cn/rpfdb/.

Acknowledgements

We would like to thank all team members for their support and RPFdb users for their invaluable feedback and suggestions.

Author contributions: H.W. and Z.X.: conceptualization and writing; Y.W.: data collection, data analysis, writing and website redesign; Y.T.: data collection, data analysis and data presentation.

Funding

National Natural Science Foundation of China [32270700 to H.W.W., in part]; Guangdong Basic and Applied Basic Research Foundation [2024A1515010103 to H.W.W.]; Guangzhou Science and Technology Program key projects [2024A03J0158 to H.W.W.]. Funding for open access charge: National Natural Science Foundation of China.

Conflict of interest statement

None declared.

References

1. Schwanhausser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W. and Selbach,M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.
2. Jia,X.C., He,X.Y., Huang,C.T., Li,J., Dong,Z.G. and Liu,K.D. (2024) Protein translation: biological processes and therapeutic strategies for human diseases. *Signal Transduct. Target. Ther.*, **9**, 44.
3. Hershey,J.W., Sonenberg,N. and Mathews,M.B. (2012) Principles of translational control: an overview. *Cold Spring Harb. Perspect. Biol.*, **4**, a011528.
4. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R. and Weissman,J.S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
5. Brar,G.A. and Weissman,J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.*, **16**, 651–664.
6. Wang,H.W., Wang,Y. and Xie,Z. (2019) Computational resources for ribosome profiling: from database to web server and software. *Brief. Bioinform.*, **20**, 144–155.
7. Chen,J., Brunner,A.D., Cogan,J.Z., Nuñez,J.K., Fields,A.P., Adamson,B., Itzhak,D.N., Li,J.Y., Mann,M., Leonetti,M.D., *et al.* (2020) Pervasive functional translation of noncanonical human open reading frames. *Science*, **367**, 1140–1146.
8. Fan,X.J., Yang,Y., Chen,C.Y. and Wang,Z.F. (2022) Pervasive translation of circular RNAs driven by short IRES-like elements. *Nat. Commun.*, **13**, 3751.
9. Yang,Y., Fan,X.J., Mao,M.W., Song,X.W., Wu,P., Zhang,Y., Jin,Y.F., Yang,Y., Chen,L.L., Wang,Y., *et al.* (2017) Extensive translation of circular RNAs driven by N^6-methyladenosine. *Cell Res.*, **27**, 626–641.
10. Dever,T.E., Ivanov,I.P. and Hinnebusch,A.G. (2023) Translational regulation by uORFs and start codon selection stringency. *Genes Dev.*, **37**, 474–489.
11. Wu,Q., Wright,M., Gogol,M.M., Bradford,W.D., Zhang,N. and Bazzini,A.A. (2020) Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.*, **39**, e104763.
12. Collart,M.A. and Weiss,B. (2020) Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Res.*, **48**, 1043–1055.
13. Liu,Y., Beyer,A. and Aebersold,R. (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell*, **165**, 535–550.
14. Walsh,D., Mathews,M.B. and Mohr,I. (2013) Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harb. Perspect. Biol.*, **5**, a012351.
15. Wang,H.W., Yang,L.D., Wang,Y., Chen,L.S., Li,H.H. and Xie,Z. (2019) RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res.*, **47**, D230–D234.
16. Kiniry,S.J., Judge,C.E., Michel,A.M. and Baranov,P.V. (2021) Trips-Viz: an environment for the analysis of public and user-generated ribosome profiling data. *Nucleic Acids Res.*, **49**, W662–W670.
17. Michel,A.M., Kiniry,S.J., O'Connor,P.B.F., Mullan,J.P. and Baranov,P.V. (2018) GWIPS-viz: 2018 update. *Nucleic Acids Res.*, **46**, D823–D830.
18. Liu,W.T., Xiang,L.P., Zheng,T.K., Jin,J.J. and Zhang,G. (2018) TranslatomeDB: a comprehensive database and cloud-based analysis platform for translatome sequencing data. *Nucleic Acids Res.*, **46**, D206–D212.
19. Liu,Q., Peng,X., Shen,M.Y., Qian,Q., Xing,J.L., Li,C. and Gregory,R.I. (2023) Ribo-uORF: a comprehensive data resource of upstream open reading frames (uORFs) based on ribosome profiling. *Nucleic Acids Res.*, **51**, D248–D261.
20. Manske,F., Ogoniak,L., Jürgens,L., Grundmann,N., Makalowski,W. and Wethmar,K. (2023) The new uORFdb:

integrating literature, sequence, and variation data in a central hub for uORF research. *Nucleic Acids Res.*, **51**, D328–D336.
21. Li,H.H., Xie,M.Z., Wang,Y., Yang,L.D., Xie,Z. and Wang,H.W. (2021) riboCIRC: a comprehensive database of translatable circRNAs. *Genome Biol.*, **22**, 79.
22. Olexiouk,V., Van Criekinge,W. and Menschaert,G. (2018) An update on sORFs.Org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.*, **46**, D497–D502.
23. Xie,S.Q., Nie,P., Wang,Y., Wang,H.W., Li,H.Y., Yang,Z.L., Liu,Y.Z., Ren,J. and Xie,Z. (2016) RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res.*, **44**, D254–D258.
24. Ji,Z., Song,R., Regev,A. and Struhl,K. (2015) Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife*, **4**, e08890.
25. Ji,Z. (2018) RibORF: identifying genome-wide translated open reading frames using ribosome profiling. *Curr. Protoc. Mol. Biol.*, **124**, e67.
26. Mudge,J.M., Ruiz-Orera,J., Prensner,J.R., Brunet,M.A., Calvet,F., Jungreis,I., Gonzalez,J.M., Magrane,M., Martinez,T.F., Schulz,J.F., *et al.* (2022) Standardized annotation of translated open reading frames. *Nat. Biotechnol.*, **40**, 994–999.
27. Leppek,K., Byeon,G.W., Kladwang,W., Wayment-Steele,H.K., Kerr,C.H., Xu,A.D.F., Kim,D., Topkar,V.V., Choe,C., Rothschild,D., *et al.* (2022) Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat. Commun.*, **13**, 1536.
28. Li,F.J., Xing,X.D., Xiao,Z.T., Xu,G. and Yang,X.R. (2020) RiboMiner: a toolset for mining multi-dimensional features of the translatome with ribosome profiling data. *BMC Bioinformatics*, **21**, 340.
29. VanInsberghe,M., van den Berg,J., Andersson-Rolf,A., Clevers,H. and van Oudenaarden,A. (2021) Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. *Nature*, **597**, 561–565.
30. Ozadam,H., Tonn,T., Han,C.M., Segura,A., Hoskins,I., Rao,S., Ghatpande,V., Tran,D., Catoe,D., Salit,M., *et al.* (2023) Single-cell quantification of ribosome occupancy in early mouse development. *Nature*, **618**, 1057–1064.
31. Zeng,H., Huang,J., Ren,J., Wang,C.K., Tang,Z., Zhou,H., Zhou,Y., Shi,H., Aditham,A., Sui,X., *et al.* (2023) Spatially resolved single-cell translatomics at molecular resolution. *Science*, **380**, eadd3067.

Received: July 22, 2024. **Revised:** August 15, 2024. **Editorial Decision:** September 2, 2024. **Accepted:** September 5, 2024