Article

Single-cell Ribo-seq reveals cell cycle-dependent translational pausing

https://doi.org/10.1038/s41586-021-03887-4

Received: 26 November 2020

Accepted: 6 August 2021

Published online: 8 September 2021

Check for updates

Michael VanInsberghe[1]✉, Jeroen van den Berg[1], Amanda Andersson-Rolf[1], Hans Clevers[1] & Alexander van Oudenaarden[1]✉

Single-cell sequencing methods have enabled in-depth analysis of the diversity of cell types and cell states in a wide range of organisms. These tools focus predominantly on sequencing the genomes[1], epigenomes[2] and transcriptomes[3] of single cells. However, despite recent progress in detecting proteins by mass spectrometry with single-cell resolution[4], it remains a major challenge to measure translation in individual cells. Here, building on existing protocols[5–7], we have substantially increased the sensitivity of these assays to enable ribosome profiling in single cells. Integrated with a machine learning approach, this technology achieves single-codon resolution. We validate this method by demonstrating that limitation for a particular amino acid causes ribosome pausing at a subset of the codons encoding the amino acid. Of note, this pausing is only observed in a sub-population of cells correlating to its cell cycle state. We further expand on this phenomenon in non-limiting conditions and detect pronounced GAA pausing during mitosis. Finally, we demonstrate the applicability of this technique to rare primary enteroendocrine cells. This technology provides a first step towards determining the contribution of the translational process to the remarkable diversity between seemingly identical cells.

Single-cell ribosome sequencing (scRibo-seq) combines nuclease footprinting with small-RNA library construction and a size enrichment to measure translation dynamics in single cells (Fig. 1a). By directly integrating this process into a one-pot reaction, we were able to markedly increase the sensitivity and scalability of existing ribosome profiling techniques. In brief, single live cells are sorted into a lysis buffer containing cycloheximide to stabilize and halt ribosomes on transcripts. Exposed RNA is then digested by micrococcal nuclease (MNase) and the resulting ribosome-protected footprints (RPFs) are released. These footprints are converted into sequencing libraries by ligating adaptors that contain a unique molecular identifier (UMI) and priming sites for subsequent cDNA synthesis and indexing PCR. Finally, the reaction products are pooled and size-selected to enrich for inserts that correspond to the typical RPF length.

To validate this method, we generated scRibo-seq libraries from HEK 293T and hTERT RPE-1 cells. The resulting single-cell libraries detect 3,348 ± 15 genes with 10,451 ± 85 unique reads per cell (mean ± s.e.m.) (Extended Data Fig. 1) and exhibit several features that are characteristic of ribosomal profiling experiments (Fig. 1b, c, Extended Data Fig. 2). First, the fragments map predominantly to coding sequences (CDSs) (Fig. 1b, c), with their 5′ ends sharply increasing approximately 15 nucleotides (nt) upstream of the start codon and decreasing approximately 18 nt upstream of the stop codon (Fig. 1b, left and right). Second, there is an increase in local density over both the start and stop codons (Fig. 1b), originating from ribosomes that are in the initiation and termination phases of translation. Third, the 5′ end of the fragments shows a clear but modest 3-nt periodicity along the CDS (Fig. 1b), with 40.7% ± 0.07% of the 5′ ends of the footprints occurring in frame 1 (Extended Data Fig. 2c). Finally, the mapping frequencies to common contaminants, different biotypes and across the untranslated regions (UTRs) and CDS of protein-coding genes are all similar to those from conventional ribosome profiling methods[6,8–10] (Extended Data Fig. 2).

scRibo-seq libraries also display patterns associated with the MNase digestion. Consistent with previous reports[6,11], we observe a broad distribution of footprint lengths, a complex association between fragment length and the predominant frame of the 5′ end (Extended Data Fig. 2c, top row), and a strong preference for an MNase cut to occur to the 5′ of an adenine or uracil (Extended Data Fig. 3a). In conventional ribosome profiling, the exit (E), peptidyl (P) and aminoacyl (A) active sites of the ribosome can be positioned in the footprint using a constant offset from the end of the read (Extended Data Fig. 3b), thereby positioning ribosomes on transcripts with single-codon resolution. We predicted that the strong sequence bias of MNase would result in incomplete digestion of the RPFs, resulting in a sequence-dependent relationship between the 5′ end of the fragment and the active sites.

We trained a random forest classifier to correct the MNase sequence bias. Similar to previous approaches[12], our model predicts the offset between the 5′ end of the footprint and the ribosome A site given the length of the fragment and the sequence context around the 5′ and 3′ cut sites (Extended Data Fig. 3c). The classifier was trained on reads that span a stop codon, achieving a high prediction accuracy (96.5% ± 0.06%, fivefold cross-validation) (Extended Data Fig. 3d). The accuracy was further confirmed by examining footprints within the CDS, where 63.9% ± 0.07% of predicted A sites were found to be in frame (Extended

[1]Oncode Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Utrecht, The Netherlands.
✉e-mail: m.vaninsberghe@hubrecht.eu; a.vanoudenaarden@hubrecht.eu

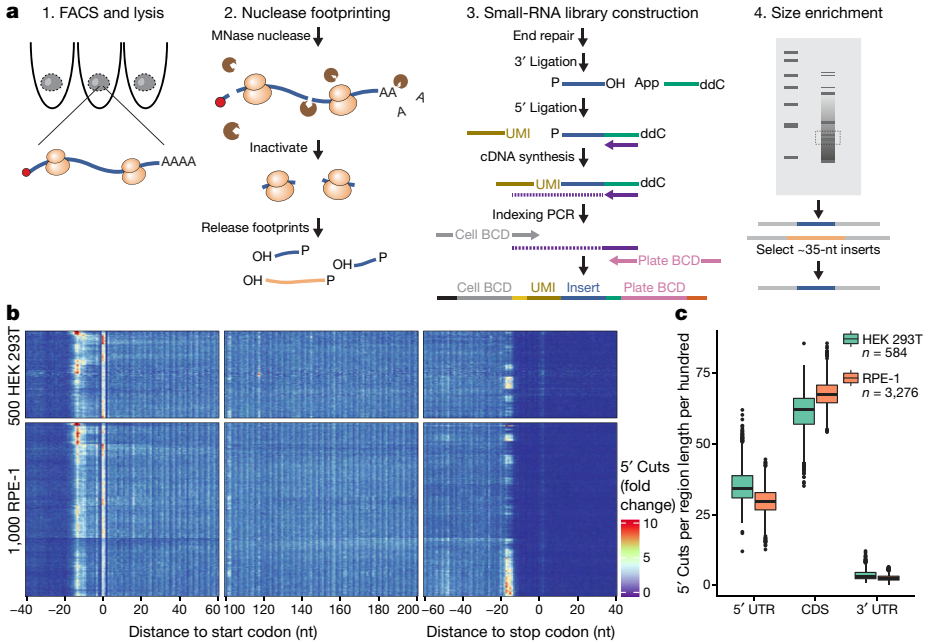


Fig. 1 | scRibo-seq measures translation in single cells. a, Schematic of the scRibo-seq workflow. BCD, barcode; FACS, fluorescence-activated cell sorting. **b**, Heat map of the fold change of the number of ribosome footprints aligning along metagene regions around the start codon (left), in the CDS (middle) and around the stop codon (right). The mapping coordinate of the 5′ end of each read is shown. **c**, Region-length normalized distributions of RPF mapping frequencies in the 5′ UTR, CDS and 3′ UTR regions of protein-coding transcripts. In the box plots, the middle line indicates the median, the box delimits the first and third quartiles and the whiskers show the range.

Data Figs. 2c, 3g); this result is reproducible between cells (Extended Data Fig. 3f), and is again similar to that obtained by conventional ribosome profiling methods (RPE-1 (ref.[10]): 61.4% ± 4.2% of 29-nt reads; HEK 293T[8]: 75.4% ± 1.4% of 29-nt reads; and HEK 293T[9]: 55.4% ± 17.5% of 28-nt reads) (Extended Data Fig. 2c). As expected, the sequence composition around the 5′ end has the highest permutation importance, followed by the fragment length, and only a minor contribution from the 3′ sequence context (Extended Data Fig. 3e), suggesting that our model is indeed capturing the MNase sequence bias.

Ribosomes have previously been observed to dwell over a subset of codons encoding essential amino acids that have been removed from the culture medium[6,13]. Ribosome profiling exposes this pausing as an increase in footprint density over the affected codons. To further validate that scRibo-seq measures translation dynamics, we removed arginine and leucine from HEK 293T culture media for 3 and 6 h before sorting. By comparing the change in codon occupancy in the predicted E, P and A sites between pseudobulk analyses of the depletion and rich conditions, we observed treatment-specific pausing (Fig. 2a). For example, arginine depletion resulted in footprints more frequently containing CGC and CGU codons compared to rich media (Fig. 2a, blue), and this increase was not seen upon removal of leucine (Fig. 2a, green). Similarly, an increase in UUA occupancy was seen only with leucine starvation.

Treatment-specific pausing was also evident in single cells. Reiterating our previous findings from the pseudobulk analysis, we again observed that pausing on arginine and leucine codons was seen only in cells isolated from the starvation conditions, and only over a subset of codons encoding the removed amino acids (Fig. 2b, Extended Data Fig. 4a). Furthermore, the position in the ribosome footprint where pausing occurs was roughly as expected, as the increases in codon occupancies were only apparent in and upstream of the A site. This ribosome-site specificity was also apparent on several codons that have been previously associated with ribosome pausing, with, for example, AAA and GAA showing increased occupancies in the A sites[14,15], and proline codons in the E sites[16] (Extended Data Fig. 4a).

Of note, only a subset of the cells from each limitation condition exhibited a pausing response (41 out of 155 from arginine limitation and 24 out of 202 from leucine limitation). Clustering cells using the RPF counts per CDS identified four clusters distinguished by cell cycle marker genes (Fig. 2c, f, Extended Data Fig. 4e). On the basis of these clusters, it is apparent that the cell cycle state has an influence on the effect of amino acid limitation on translational pausing. The vast majority (89.7%) of cells that paused under arginine limitation were in either early (cluster 2, 5 cells) or late (cluster 1, 30 cells) S phase (CGC: $P = 0.0096$, CGU: $P = 0.014$, Fisher's exact test) (Extended Data Fig. 4b–d), whereas the cells that respond to leucine limitation are not significantly associated with any cluster (UUA: $P = 0.33$, Fisher's exact test) (Fig. 2d, Extended Data Fig. 4d).

The location of ribosome pausing on individual genes was also evident in single cells. Examining the RPF density over *H3C2*, one of the genes that exhibits an increase in CGC pausing under arginine starvation, revealed several pausing hotspots (Fig. 2e, g). The most prominent of these regions includes two successive CGC codons (Fig. 2e), potentially explaining the increased density at this location compared with other identical codons on this transcript. Additionally, these repetitive codons may cause the increase in CGC and CGU occupancy downstream of the A and P sites seen in Fig. 2b.

After seeing that the cell cycle state can affect the response to amino acid limitation, we next tested whether translational properties changed through the unperturbed cell cycle. Translational regulation has previously been identified as an important cell cycle control mechanism[10,17]. However, these studies only coarsely resolve the main cell cycle states and rely on arresting or synchronizing cells with methods that also act on translational machinery[18–20]. We generated scRibo-seq libraries from 3,276 single hTERT-RPE-1 cells expressing fluorescent ubiquitination-based cell cycle indicators (FUCCI)[21] collected from interphase (2,261 cells), contact-inhibition G0 (291 cells), and mitotic shake-off (724 cells) fractions (Fig. 3a, f). Clustering single cells using the RPF counts per CDS identifies eight clusters delineating the main phases of the cell cycle (Fig. 3b). The progression and identity of these

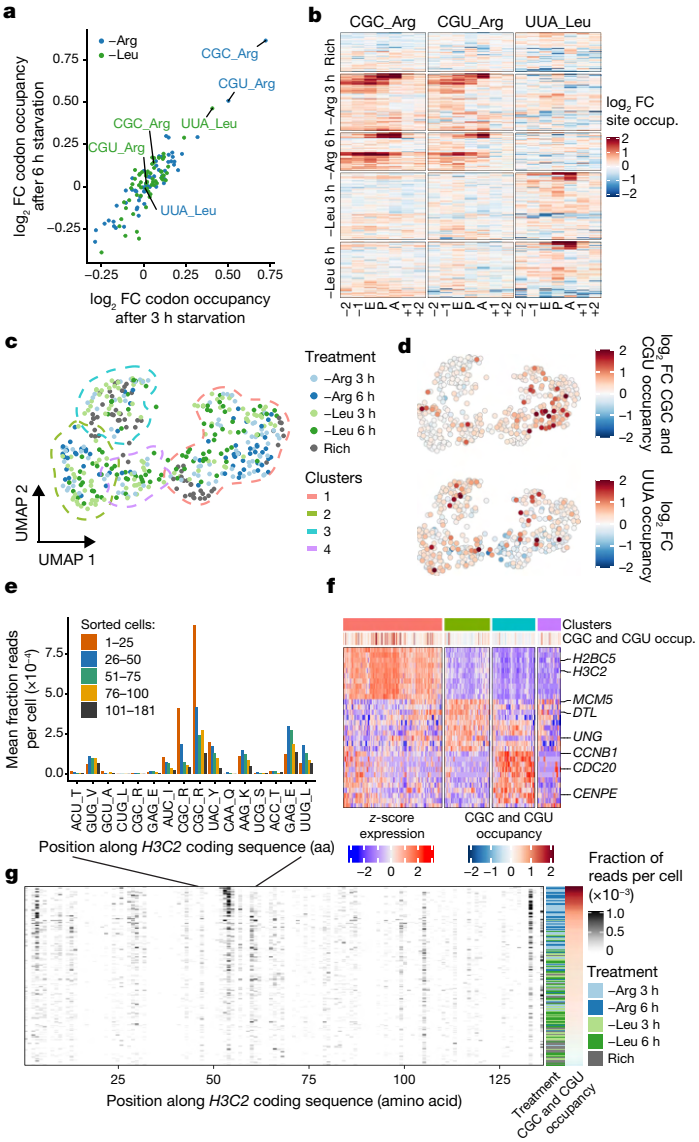


Fig. 2 | Ribosome pausing under amino acid limitation. a, Codon occupancy in ribosome exit (E), peptidyl (P) and aminoacyl (A) sites from reads pooled from all single cells from each condition. FC, fold change. **b**, Heat map of the fold change in codon occupancy in sites relative to the ribosome active sites. **c**, UMAP ($n = 421$ cells) of the single-cell RPF libraries showing derived clusters (dashed outline) and limitation condition (point colour). UMAP and clusters were determined using RPF counts per CDS. **d**, UMAP analysis showing the mean $\log_2$ fold change in codon occupancy for arginine- (top) and leucine- (bottom) limitation conditions. **e**, Average of P-site occupancy along a section of *H3C2* for cells sorted and grouped on the basis of their global arginine pausing. **f**, Heat map showing RPF counts per CDS of the top marker genes for each cell cluster. Occupancy colour scales as in **d**. **g**, Heat map showing the single-cell P-site occupancy along *H3C2*. Cells are sorted by mean CGC and CGU occupancy.

clusters closely follow those expected based on fluorescence measurements of the FUCCI markers collected during index sorting (Fig. 3d, e, g). Pseudotime ordering further resolves this progression through the cell cycle, establishing trajectories through the uniform manifold approximation and projection (UMAP) (Fig. 3c) and FUCCI markers (Fig. 3h), and revealing the translation dynamics of 1,853 significantly differentially translated genes (Extended Data Fig. 5a). Additionally, the change in abundance of several canonical cell cycle markers follows the expected pattern, further confirming the cell ordering.

In addition to this expected fluctuation in the RPF abundance of numerous genes over the cell cycle, the frequency of certain codons in the ribosome footprints also varies. While most codons have constant frequencies of occurrence across ribosome sites and cell cycle stages (for example, CAG, Fig. 3i), we identified ten codons whose frequencies of occurrence in at least one of the ribosome active sites significantly changed throughout the cell cycle (adjusted P (P_{adj}) $< 10^{-15}$ Wilcoxon rank-sum test) (Extended Data Fig. 5b).

Most of these variable codons display similar changes in occupancy in not only the ribosome E, P and A sites but also in positions immediately upstream (−1 and −2) and downstream (+1 and +2). For example, UGC is approximately 1.4 times more likely to occur in all RPF sites in cells in G0 and late G1 (clusters 3 and 8; 1.06% ± 0.003% of RPF sites) than in cells in mitosis (cluster 7; 0.78% ± 0.002% of RPF sites; $P < 2 \times 10^{-16}$, Wilcoxon rank-sum test) (Fig. 3i). However, because these changes are not isolated to specific ribosome active sites, they are probably the result of fluctuations in codon usage[22] rather than changes to translational elongation processes.

Notably, CGC and CGU, the two codons that show the strongest response to arginine limitation in HEK 293T cells, also show these site-agnostic increases in RPE-1 cells in late S phase (cluster 5; Fig. 3i,

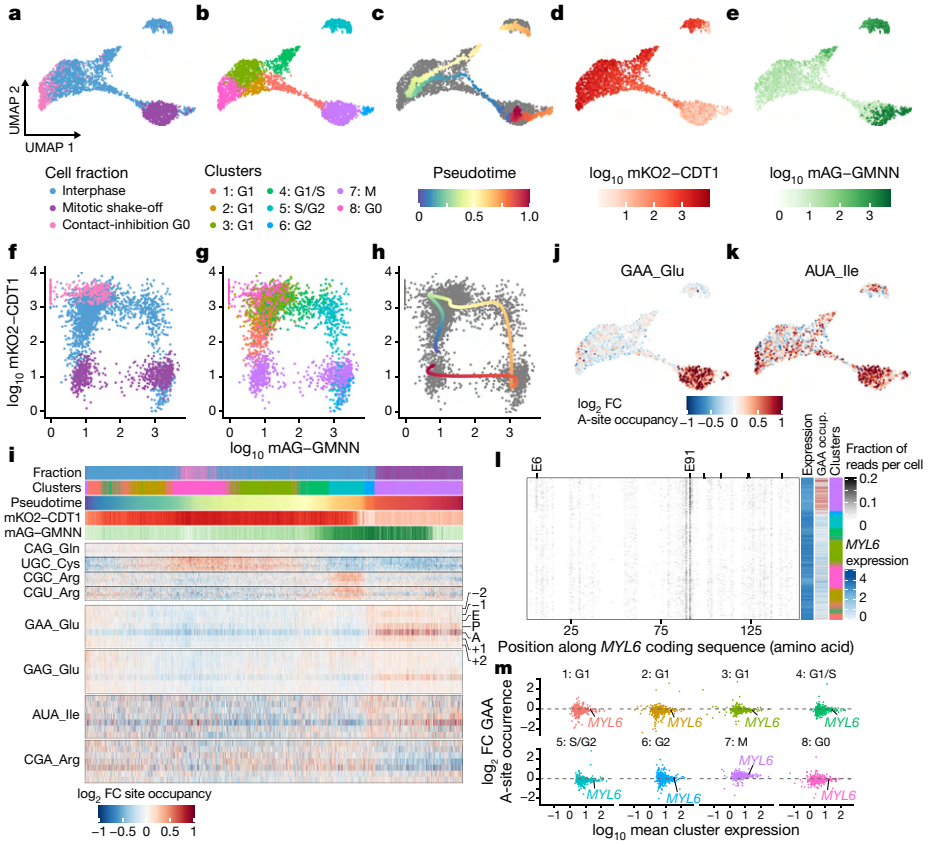


Fig. 3 | Ribosome pausing during the cell cycle. a–**e**, UMAPs ($n = 3,276$ cells) determined using RPF counts per CDS illustrating cell fractions (**a**), cell clusters (**b**), pseudotime trajectory (**c**), monomeric Kusabira-Orange 2 (mKO2)–CDT1 marker fluorescence (**d**) and monomeric Azami-Green (mAG)–GMNN FUCCI marker fluorescence (**e**). **f**–**h**, Scatter plots of the FUCCI markers ($n = 3,276$ cells) with colours denoting cell fractions (**f**), cell clusters (**g**) and pseudotime trajectory (**h**) as in **a**–**c**. **i**, Heat map showing ribosome-site-specific pausing over example codons in single cells ordered on the basis of cell cycle progression. **j**, **k**, UMAP analysis showing GAA pausing (**j**) and AUA pausing (**k**). **l**, Heat map showing the distribution of RPF A sites along the *MYL6* CDS. Cells are ordered on the basis of cell cycle progression. GAA codons are denoted by ticks along the top of the graph. Occup., occupancy. **m**, Scatter plots showing the fold change in gene-wise A-site frequency of occurrence for each cell cluster over the background.

Extended Data Fig. 6b). A substantial fraction of the translational activity in these cells goes towards producing histones (6.8% ± 0.1% of RPFs per cell), which are enriched for CGC and CGU codons (Extended Data Fig. 6a). Although these observations again suggest that these site-agnostic increases reflect a change in codon usage, the increased demand for arginine may indicate why HEK 293T cells in S phase are more susceptible to arginine limitation (Fig. 2a, b).

Conversely, the other codons exhibit site-specific changes in cells undergoing mitosis. Among the codons with variable frequencies of occurrence along the cell cycle are four whose A-site occupancies either increase (GAA, GAG and AUA) or decrease (CGA) in mitotic cells, whereas the other active sites remain constant (Fig. 3i; mitotic cells: cluster 7, purple). Of these, the increase in A-site pausing over GAA is the most pronounced and stage-specific (Fig. 3i, j), with 5.6% ± 0.07% of the RPFs from cells in mitosis containing a GAA in the A site, compared with 3.8% ± 0.01% in the other stages ($P < 2 \times 10^{-16}$, Wilcoxon rank-sum test). Not all codons follow this same trend, however. For example, cells that are in late mitosis (marked by the sharp decrease in monomeric Azami-Green (mAG–GMNN) fluorescence) exhibit higher AUA pausing than those in early mitosis (Fig. 3i, k), whereas CGA pausing decreases in mitotic and G0 cells compared with the other stages (Fig. 3i). These stage-specific pausing signals are distinct from changes in codon usage as they are specifically isolated to ribosome A sites.

These changes in A-site pausing are global, affecting the majority of translated genes. Comparing the gene-wise frequency of occurrence of GAA codons in RPF A sites between each cluster and the background reveals that most genes experience increased GAA pausing during mitosis (Fig. 3m). For example, in mitotic cells 22.7% ± 0.5% of the RPFs aligning to *MYL6* have a GAA in the A site, with most of these occurring at E6 and E91; in the other stages, only 15.4% ± 0.2% of the A sites contain a GAA (Fig. 3l). Averaged across all genes, this is a modest increase of 1.23 ± 0.01-fold; however, it is widespread, as 41.5% of GAA-containing genes detected across more than three clusters (165 of 398 detected genes) show a significant increase in A-site GAA pausing in mitotic cells. While not as strong, this same trend is also observed for GAG, AUA and CGA (Extended Data Fig. 7). These global stage-specific changes to A-site pausing may reflect global alterations in translation elongation dynamics during mitosis.

Having demonstrated scRibo-seq on cell lines, we next generated ribosome profiles for primary mouse intestinal enteroendocrine (EEC) cells, a population of rare cells in the gastrointestinal epithelium (less than 1% of cells) that produces and secretes diverse hormones in response to nutrient stimuli[23]. They are further subclassified on the basis of the hormones that they produce, with the seven cell lineages producing different hormones as they mature, resulting in up to 20 different EEC cell types being described[24,25]. Their scarcity, diversity and plasticity make primary EEC cells inaccessible to existing ribosome profiling methods, making it challenging to study post-transcriptional and translational regulation of their behaviours. We generated ribosomal profiles from 350 single mouse EEC cells expressing a bi-fluorescent *Neurog3* reporter[24] that were isolated from intestinal crypts (Extended Data Figs. 8, 9). Clustering cells on the basis of the RPF counts per CDS

identifies eight clusters representing the main EEC cell types in the crypts that are delineated by the translation of established hormone marker genes (Extended Data Figs. 8a, 9a, b). Among the cells are two minority subpopulations that show genome-wide ribosome pausing over CAG-glutamine ($n = 16$ cells) and GAA-glutamic acid ($n = 6$) codons (Extended Data Figs. 8f–k, 9c). Notably, the GAA-pausing population is only present in the late enterochromaffin cluster (6 out of 29 cells), whereas the CAG-pausing cells were distributed between the cell clusters (GAA: $P = 1.9 \times 10^{-7}$; CAG: $P = 0.014$; Fisher's exact test). Together, these results establish that scRibo-seq is directly applicable to complex primary samples, enabling the measurement of translational dynamics in rare cell populations.

scRibo-seq measures translation at the single-cell level, filling a crucial gap in existing capabilities for single-cell genomics. Our results demonstrate that scRibo-seq provides a marker-free and transgene-free method for ribosomal profiling with the sensitivity and resolution to measure ribosome behaviour down to individual codons on specific transcripts in populations of single cells. Compared with the recently described Ribo-STAMP[26], which uses APOBEC-mediated RNA editing to identify transcripts that have been associated with ribosomes, scRibo-seq provides an instantaneous snapshot of translation, has single-codon resolution, and does not require the expression of an exogenous fusion protein. These unique capabilities enabled us to examine translation during the mammalian cell cycle in detail, providing evidence supporting widespread changes to translational regulation during mitosis. Furthermore, our application of scRibo-seq to primary EEC cells demonstrates its applicability to rare primary samples, enabling the systematic exploration of translation in samples that would be impossible to access with bulk assays. We anticipate that this method will see broad application, particularly in highly dynamic systems such as development, in which rare and short-lived populations are impossible to measure with existing techniques.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03887-4.

1. Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
2. Smallwood, S. A. et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
3. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
4. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
5. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
6. Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Mol. Cell* **71**, 229–243.e11 (2018).
7. Reid, D. W., Shenolikar, S. & Nicchitta, C. V. Simple and inexpensive ribosome profiling analysis of mRNA translation. *Methods* **91**, 69–74 (2015).
8. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).
9. Martinez, T. F. et al. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).
10. Tanenbaum, M. E., Stern-Ginossar, N., Weissman, J. S. & Vale, R. D. Regulation of mRNA translation during mitosis. *eLife* **4**, e07957 (2015).
11. Gerashchenko, M. V. & Gladyshev, V. N. Ribonuclease selection for ribosome profiling. *Nucleic Acids Res.* **45**, e6 (2017).
12. Fang, H. et al. Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. *Cell Syst.* **6**, 180–191.e4 (2018).
13. Subramaniam, A. R., Pan, T. & Cluzel, P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl Acad. Sci. USA* **110**, 2419–2424 (2013).
14. Zinshteyn, B. & Gilbert, W. V. Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* **9**, e1003675 (2013).
15. Nedialkova, D. D. & Leidel, S. A. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* **161**, 1606–1618 (2015).
16. Artieri, C. G. & Fraser, H. B. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* **24**, 2011–2021 (2014).
17. Stumpf, C. R., Moreno, M. V., Olshen, A. B., Taylor, B. S. & Ruggero, D. The translational landscape of the mammalian cell cycle. *Mol. Cell* **52**, 574–582 (2013).
18. Coldwell, M. J. et al. Phosphorylation of eIF4GII and 4E-BP1 in response to nocodazole treatment: a reappraisal of translation initiation during mitosis. *Cell Cycle* **12**, 3615–3628 (2013).
19. Ly, T., Endo, A. & Lamond, A. I. Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *elife* **4**, e04534 (2015).
20. Miettinen, T. P., Kang, J. H., Yang, L. F. & Manalis, S. R. Mammalian cell growth dynamics in mitosis. *elife* **8**, e44700 (2019).
21. Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
22. Frenkel-Morgenstern, M. et al. Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.* **8**, 572 (2012).
23. Gribble, F. M. & Reimann, F. Enteroendocrine cells: chemosensors in the intestinal epithelium. *Annu. Rev. Physiol.* **78**, 277–299 (2016).
24. Gehart, H. et al. Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. *Cell* **176**, 1158–1173.e16 (2019).
25. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
26. Brannan, K. W. et al. Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. *Nat. Methods* **18**, 507–519 (2021).

Methods

Cell culture and dissociation

HEK 293T cells were obtained from the Medema laboratory (R.H.M., Department of Cell Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands) and were cultured in DMEM (Gibco) supplemented with 10% FBS (Gibco), 1× GlutaMAX (Gibco) and 1× Pen-Strep (Gibco) at 37 °C and 5% CO_2. HEK 293T cells routinely tested negative for *Mycoplasma* contamination and were not authenticated. For amino acid limitation experiments, HEK 293T cells were cultured to about 70% confluency in 'rich' medium based on powdered DMEM medium for stable isotope labelling using amino acids in cell culture (SILAC) (Thermo Fisher Scientific) that was supplemented with 10% dialysed FBS (Thermo Fisher Scientific), 105 mg l^{-1} L-leucine (Sigma-Aldrich), 84 mg l^{-1} L-arginine HCl (Sigma-Aldrich) and 146 mg l^{-1} L-lysine HCl (Sigma-Aldrich). Three and six hours before sorting, cells were washed once with phosphate buffered saline (PBS) and resuspended in medium that did not contain either arginine or lysine. Before sorting, cells were mechanically dissociated to a single-cell suspension by pipetting up and down, washed and resuspended in PBS containing DAPI (Thermo Fisher Scientific) as a viability stain and bovine serum albumen (BSA, Thermo Fisher Scientific) to reduce aggregation. Cells were passed through a 20-µm mesh before sorting and all viable single cells were sorted (Extended Data Fig. 10a).

RPE-1 hTERT FUCCI cells were obtained from the Medema lab and were cultured in DMEM supplemented with 10% FBS (Gibco), 1× GlutaMAX (Gibco) and 1× Pen-Strep (Gibco) at 37 °C with 5% CO_2. RPE-1 hTERT FUCCI cells routinely tested negative for *Mycoplasma* contamination and were not authenticated. For the RPE-1 cell cycle experiments, we used previously characterized RPE-1 hTERT FUCCI cells[27], and generated three fractions: interphase, mitotic shake-off and G0-arrested. For the interphase fraction, 7.5×10^4 cells were plated in a 6-well plate and collected by trypsinization (TrypLE, Gibco) 36 h later. For the mitotic fraction, 3×10^6 cells were plated in a 145-mm dish and were collected 36 h later by gently tapping the culture dish and collecting the medium (otherwise known as a mitotic shake-off). Finally, for the G0-arrested fraction, 1×10^5 cells were plated in a 24-well plate and collected 72 h later by trypsinization. Cells were washed once and resuspended in PBS containing DAPI as a viability stain and BSA to reduce aggregation. Cells were passed through a 20-µm mesh before sorting. For the interphase fraction, viable single cells were sorted (Extended Data Fig. 10b). For the G0 fraction, viable single mAG-negative and mKO2-positive or -high cells were sorted (Extended Data Fig. 10c). To further enrich the mitotic shake-off, we sorted viable, mKO2-negative cells (Extended Data Fig. 10d).

Mouse EEC cells were isolated from the intestines of Neurog3Chrono mice, closely following the methods outlined by Gehart et al.[24]. In brief, mouse small intestines were collected, cleaned, flushed with PBS0, and separated into proximal, medial and distal sections. Pieces were cut open and villi were scraped off with a glass cover slip and discarded. Tissue pieces were then washed in cold PBS0 before transferring to PBS0 with 2 mM EDTA (Gibco), incubated at 4 °C for 30 min on a roller, and then vigorously shaken. Detached crypts were pelleted, resuspended in warm TrypLE Select (Gibco), and mechanically disrupted by pipetting to generate single-cell suspensions. Single-cell suspensions were washed twice in Advanced DMEM/F12 (Gibco), strained with a 20-µm mesh, and resuspended in Advanced DMEM/F12 containing 4 mM EDTA and 1 µg ml^{-1} DAPI for sorting. Example gating strategies are shown in Extended Data Fig. 10e.

Mice

All mouse experiments were conducted under the project license AVD8010020151 granted by the Dier Experiment Commissie/Animal Experimentation Committee (DEC) or Central Committee Animal Experimentation (CCD) of the Dutch government and approved by the Hubrecht Institute Animal Welfare Body (IvD). The Neurog3Chrono allele was maintained on a mixed *Mus musculus* C57BL/6 background. Animals used in the experiments were between 8 and 22 weeks of age. Both male and female mice were used for the experiments. Mice were housed in open housing with 14:10 h light:dark cycle at 24 °C and 45–70% relative humidity with food and water ad libitum. The intestines from two individuals were pooled together during cell dissociation; randomization and blinding were not performed.

FACS

Following dissociation, HEK 293T and RPE-1 cells were washed once in 1× PBS0, resuspended in PBS0 with 0.1% BSA (Thermo Fisher) and 1 µg ml^{-1} DAPI, and passed through a 20-µm mesh. Single cells were index sorted using a BD FACS Influx with the following settings: sort objective single cells, a drop envelope of 1.0 drop, a phase mask of 10/16, extra coincidence bits of maximum 16, drop frequency of 38 kHz, a nozzle of 100 µm with 18 PSI and a flowrate of approximately 100 events per second, which results in a minimum sorting time of approximately 5 min per plate.

Doublets, debris, and dead cells were excluded by gating forward and side scatter in combination with the DAPI channel. For the hTERT RPE-1 FUCCI cells, the measurements in the monomeric Azami-Green (mAG) and monomeric Kusabira-Orange 2 (mKO2) channels were used in combination with the cell preparation treatments to enrich G0 and mitotic populations. For the mouse intestinal EEC cells, the measurements of dTomato and mNeonGreen were used to select EEC cells expressing the Neurog3Chrono reporter[24] and DAPI was used to exclude dead cells. Fluorescence intensities from all channels were stored as index data.

Example gating strategies and cell frequencies for all cell fractions are shown in Extended Data Fig. 10.

Library construction

Library construction progressed through three general steps (Fig. 1a): cell lysis and ribosome footprint generation, small-RNA library preparation, and pooling and purification. Reagents were dispensed to microwell plates using either the Nanodrop II (Innovadyne Technologies) or the Mosquito (TTP Labtech). Plates were spun at 2,000*g* after each liquid transfer step.

Cell lysis and footprint digestion

Single cells were sorted into 384-well hardshell plates (Bio-Rad) that were pre-filled with 5 µl of light mineral oil (Sigma-Aldrich) and 50 nl of lysis buffer (22 mM Tris-HCl pH 7.5, 16.5 mM $MgCl_2$, 5.5 mM $CaCl_2$, 165 mM NaCl, 1.1% Triton X-100, 2.2 U $µl^{-1}$ RNaseIN Plus (Promega), 0.11 mg ml^{-1} cycloheximide (Sigma-Aldrich)). After sorting, plates were spun down at 2,000*g* for 2 min and kept on wet ice until all plates were ready for further processing. Next, 50 nl of MNase (10,500 U $µl^{-1}$, New England Biolabs) was added to each well, and plates were incubated at 37 °C for 30 min. To stop digestion, 50 nl of stop mix (0.0186 U $µl^{-1}$ Thermolabile Proteinase K (New England Biolabs), 62 mM EGTA (Sigma-Aldrich), 16.5 mM EDTA (Ambion), and 697.5 mM guanidium thiocyanate (GuSCN, Sigma-Aldrich)) was added to each well, and plates were incubated at 37 °C for 30 min then 55 °C for 10 min and held at 4 °C.

Small-RNA library preparation

After ribosome footprint digestion, libraries were constructed using a one-pot small-RNA library preparation protocol that incorporated end repair, two RNA ligations, cDNA synthesis, and an indexing PCR. First, 50 nl of end-repair mix (4.1× T4 RNA Ligase Buffer (New England Biolabs), 16.4 mM $MgCl_2$, 4.1 mM uridine triphosphate (New England Biolabs), 1.37 U $µl^{-1}$ T4 Polynucleotide Kinase (New England Biolabs) and 0.82 U $µl^{-1}$ RNaseIN Plus) was added to each well, and plates were incubated at 37 °C for 1 h and held at 4 °C. Next, 264 nl of 3′ ligation brew (1× T4 RNA Ligase Buffer (New England Biolabs), 1 µM pre-adenylated 3′ adapter (OMV630_miRNA4_3App, Supplementary Table 1; Integrated DNA Technologies), 35.5% PEG-8000 (New England Biolabs), 0.1% Tween-20 (Sigma-Aldrich), 1 U $µl^{-1}$ RNaseIN Plus, and 21.3 U $µl^{-1}$ T4 RNA Ligase 2 Truncated KQ (New England Biolabs)) was added to

each well and plates were incubated at 4 °C for 18 h. The cDNA synthesis primer was then pre-annealed to the 3′ ligation products by adding 50 nl of the reverse transcription primer mix (5.2 μM reverse transcription primer (OMV572_miRNA4_RT, Supplementary Table 1; Integrated DNA Technologies), 13.5 μM adenosine triphosphate (ATP, New England Biolabs), and 1% Tween-20) to each well, heating to 65 °C for 1 min, 37 °C for 2 min, 25 °C for 2 min, and holding at 4 °C. Five-prime adapters were then ligated by adding 156 nl of 5′ ligation brew (1× T4 RNA Ligase Buffer, 30.75% PEG-8000, 0.1% Tween-20, 0.5 μM 5′ adapter (OMV632_miRNA5_5A_10U, Supplementary Table 1; Integrated DNA Technologies), 1.25 U μl^{-1} T4 RNA Ligase 1 (Ambion)) and incubating at 37 °C for 2 h and holding at 4 °C. Complementary DNA synthesis was then performed by adding 770 nL of reverse transcription brew (1.88× 5× RT Buffer (Thermo Fisher Scientific), 1.25 mM dNTPs (Promega), 0.1875% Tween-20, 1.875 U μl^{-1} RNaseIN Plus, and 9.375 U μl^{-1} Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific)) to each well, and heating at 50 °C for 1 h, then 85 °C for 5 min and holding at 4 °C. Finally, single-cell libraries were indexed during PCR by transferring 150 nl of 20 μM unique forward index primers (miRv6-PCR_F-cbc, Supplementary Table 1; Integrated DNA Technologies) and 3.2 μl of PCR brew (1.5× Q5 Hot Start High-Fidelity 2× Master Mix (New England Biolabs), 0.15% Tween-20, and 0.94 μM reverse index primer (RPI-, Supplementary Table 1; Integrated DNA Technologies)) to each well. Plates were then incubated at 98 °C for 30 s followed by 10 cycles of 98 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s, and then a final incubation at 72 °C for 5 min and holding at 4 °C. Plates were then frozen at −20 °C until pooling.

Pooling and purification

After library construction the plates were pooled and purified. The contents of each plate were first collected in VBLOK200 reservoirs (Click Bio) by centrifuging at 2,000*g* for 2 min. The aqueous phase (~1.9 ml per plate) was separated from the light mineral oil by centrifugation, and concentrated to approximately 500 μl using *n*-butanol (Sigma-Aldrich) and diethyl ether (Sigma-Aldrich). Product was then cleaned up using AMPure XP beads (Beckman Coulter) that had been diluted 5× in bead binding buffer (20% PEG-8000 (Sigma-Aldrich) 2.5 M sodium chloride (Sigma-Aldrich)); diluted beads were added to the sample at a 2.1:1 ratio, and the final product was resuspended in 50 μl low TE buffer (LoTET, 3 mM Tris-HCl pH 8.0 (Ambion), 0.2 mM EDTA pH 8.0 (Gibco), 0.1% Tween-20). Half of each of the cleaned-up library pools was then run on a 10-cm 7% polyacrylamide gel at 200 V for ~6 h, and the ~10-bp region from 175 to 185, corresponding to an insert size of ~30–40 nt was excised. The band was then crushed and soaked overnight at 4 °C in elution buffer (5:1 LoTET:7.5 M ammonium acetate (Sigma-Aldrich)). Finally, eluate was precipitated in ethanol.

Sequencing

Libraries were sequenced using v2.5 chemistry on a NextSeq 500 (Illumina; NextSeq Control Software version 2.2.0.4; RTA version 2.4.11) with 75 cycles for read 1, 6 cycles for the i7 index read (plate index), and 10 cycles for the i5 index read (cell index).

Data analysis

Reference genomes and annotations. The reference genome and annotations were obtained from Gencode using human release 34 (GRCh38.p13) and mouse release 24 (GRCm38.p6). The reference genomes were prepared for alignment by masking all tRNA genes and pseudogenes and including unique mature tRNAs genes as artificial chromosomes. tRNA genes and pseudogenes were identified using tRNAscan-SE (version 2.0.7) using the eukaryotic model (-HQ) and the vertebrate mitochondrial model (-M vert -Q). Sequences for ribosomal RNAs were downloaded from NCBI RefSeq (human: 12S_RNR1, 16S_RNR2, RNA45SN5, RNA45SN1, RNA45SN4, RNA45SN2, RNA45SN3, RNA5S9, RNA5S1-17; mouse: Rn45s, Rn5s, 12s_16s, and Rn47s). For metagene analyses, a set of canonical transcripts was defined on the basis of the APPRIS annotations, with the longer isoforms being selected in cases of multiple primary isoforms.

Read processing. Reads were first demultiplexed using bcl2fastq (version 2.20.0.422) with --use-bases-mask Y*,I*,Y* --no-lane-splitting --mask-short-adapter-reads 0 --minimum-trimmed-read-length 0. Next, the UMI was extracted from the first 10 bases of read 1 and concatenated to the start of the cell barcode in read 2. Adapter sequences were then trimmed from read 1 using cutadapt (version 3.2) with -m 15: -a TGGAATTCTCGGGT. Trimmed reads were aligned to the reference genome using STARSolo (version 2.7.6a) with a 50-base overhang (--sjdbOverhang 50) with the following parameters: --seedSearchStartLmax 10 --alignIntronMax 1000000 --outFilterType BySJout --alignSJoverhangMin 8 --outFilterScoreMin 0 --outFilterMultimapNmax 1 --chimScoreSeparation 10 --chimScoreMin 20 --chimSegmentMin 15 --outFilterMismatchNmax 5. Aligned reads were deduplicated with UMI-tools (version 1.1.1) using --spliced-is-unique --per-cell --read-length --no-sort-output, and sorted using sambamba (version 0.8.0).

Read parsing and quantification. Aligned and deduplicated reads were parsed to extract the mapping coordinates and the reference sequence contexts around both the 5′ and 3′ ends of the reads. Count tables report the number of unique RPFs aligning within any annotated CDS for each gene. Reads were only counted if they aligned to the correct strand, if the trimmed length was between 30 and 45 nt long, and if the trimmed length was equal to the mapped length. Additionally, for these counts, the CDS was expanded to include 25 nt upstream and downstream of the start and stop codons.

Random forest training and prediction. A random forest model was trained to predict the A site location within an RPF read on the basis of the footprint length and the sequence context around the 5′ and 3′ ends. The model was implemented in R (version 3.6.3) using ranger (version 0.12.1) with mlr (version 2.17.1) wrappers for training, tuning, assessment, and prediction. The model was trained on reads spanning a stop codon that satisfied the counting requirements listed above. The number of nucleotides between the 5′ end of these reads and the annotated stop codon was used for training.

Single-cell expression analysis and plotting. Data are reported as mean ± s.e.m. Manipulations, statistics and plotting were performed in R (version 3.6.3) using dplyr (version 1.0.2), tidyr (version 1.1.2), ggplot2 (version 3.3.2), cowplot (version 1.1.0), and ComplexHeatmap (version 2.5.3) and Python (version 3.8.2) using numpy (version 1.20.0), pandas (version 1.2.2) and pysam (version 0.16.0).

Values for box plots were calculated using the default settings in geom_boxplot (ggplot2): the middle line represents the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the upper hinge to the largest value no further than 1.5 times the interquartile range, the lower whisker extends from the lower hinge to the smallest value no further than 1.5 times the interquartile range, and data beyond the whiskers are deemed outliers and plotted individually.

Gene set enrichment analysis was performed using FGSEA[28] (version 1.12.0) with 1×10^5 permutations, and a Benjamini–Hochberg adjustment for multiple test correction.

Analyses of single-cell RPF counts were performed using Seurat[29] (version 3.2.2) following standard workflows for normalization, batch integration, dimensionality reduction, and differential expression analysis. All clustering, differential expression, and pseudotime analyses were performed using the unique RPF counts per CDS. Differential expression testing was performed using Wilcoxon rank-sum tests between each cell cluster and all other cells. Genes were only tested if they were detected in at least 25% of the cells in either cluster, and

if there was an average of at least a log 0.25-fold change difference between the two groups (FindAllMarkers(min.pct = 0.25, logfc.threshold = 0.25)). Significance was assigned at a Bonferroni-corrected *P* value less than 0.01. To identify codons whose frequencies of occurrence varied over the cell cycle, the log-fold change threshold was decreased to 0.17 to permit smaller signals.

Cells were selected for further analysis if they met two quality control thresholds: a minimum number of unique reads aligning to protein-coding transcripts, with a minimum fraction of these reads aligning to CDS. These thresholds removed cells where the library construction and/or nuclease footprinting failed. The values of these cut-offs were adjusted slightly for each plate and sample type to compensate for differences in sequencing depth and coding-sequence coverage. Additionally, primary EEC cell doublets that were not removed during sorting were filtered using measurements of forward scatter and side scatter collected during index sorting.

Pseudotime ordering of cells along the cell cycle was done using GrandPrix[30] (version 0.1) partially refactored to work with recent versions of tensorflow (version 2.3.1) and gpflow (version 2.1.1) in Python (version 3.7.3). The model was fit with a periodic squared exponential kernel to the normalized and *z*-score transformed RPF counts and mKO2 and mAG fluorescence index data. The GrandPrix model was initialized with each cell's progression through the cell cycle as measured using the mKO2–CDT1 and mAG–GMNN index data. To compute this initial progression, 30 cells were first randomly selected from the data, organized into a path through the fluorescence space using the Concorde travelling salesman problem solver (version 03.12.19), and each cell's distance along the path was recorded. This process was repeated 500,000 times and the median path distance for each cell was used to initialize the pseudotime cell ordering for the GrandPrix model.

Reporting summary

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

Data availability

Raw sequencing data, metadata and count tables have been made available in the Gene Expression Omnibus under the accession number GSE162060. Raw sequencing data for comparisons to conventional ribosomal profiling methods were downloaded from Gene Expression Omnibus accessions GSE37744, GSE125218, GSE113751 and GSE67902.

Code availability

All scripts to process raw data and generate figures are available at https://github.com/mvanins/scRiboSeq_manuscript.

27. Shaltiel, I. A. et al. Distinct phosphatases antagonize the p53 response in different phases of the cell cycle. *Proc. Natl Acad. Sci. USA* **111**, 7313–7318 (2014).
28. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* https://doi.org/10.1101/060012 (2021).
29. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
30. Ahmed, S., Rattray, M. & Boukouvalas, A. GrandPrix: scaling up the Bayesian GPLVM for single-cell data. *Bioinformatics* **35**, 47–54 (2019).
31. Schuller, A. P. & Green, R. Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.* **19**, 526–541 (2018).

Acknowledgements This work was supported by a European Research Council Advanced grant (ERC-AdG 742225-IntScOmics) and Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP award (NWO-CW 714.016.001). This work is part of the Oncode Institute, which is partly financed by the Dutch Cancer Society. In addition, we thank the Hubrecht Sorting Facility and the Utrecht Sequencing Facility, subsidized by the University Medical Center Utrecht, the Hubrecht Institute, Utrecht University and The Netherlands X-omics Initiative (NWO project 184.034.019); H. Viñas Gaza for assistance in preparing samples; and V. Bhardwaj for discussion on data analysis.

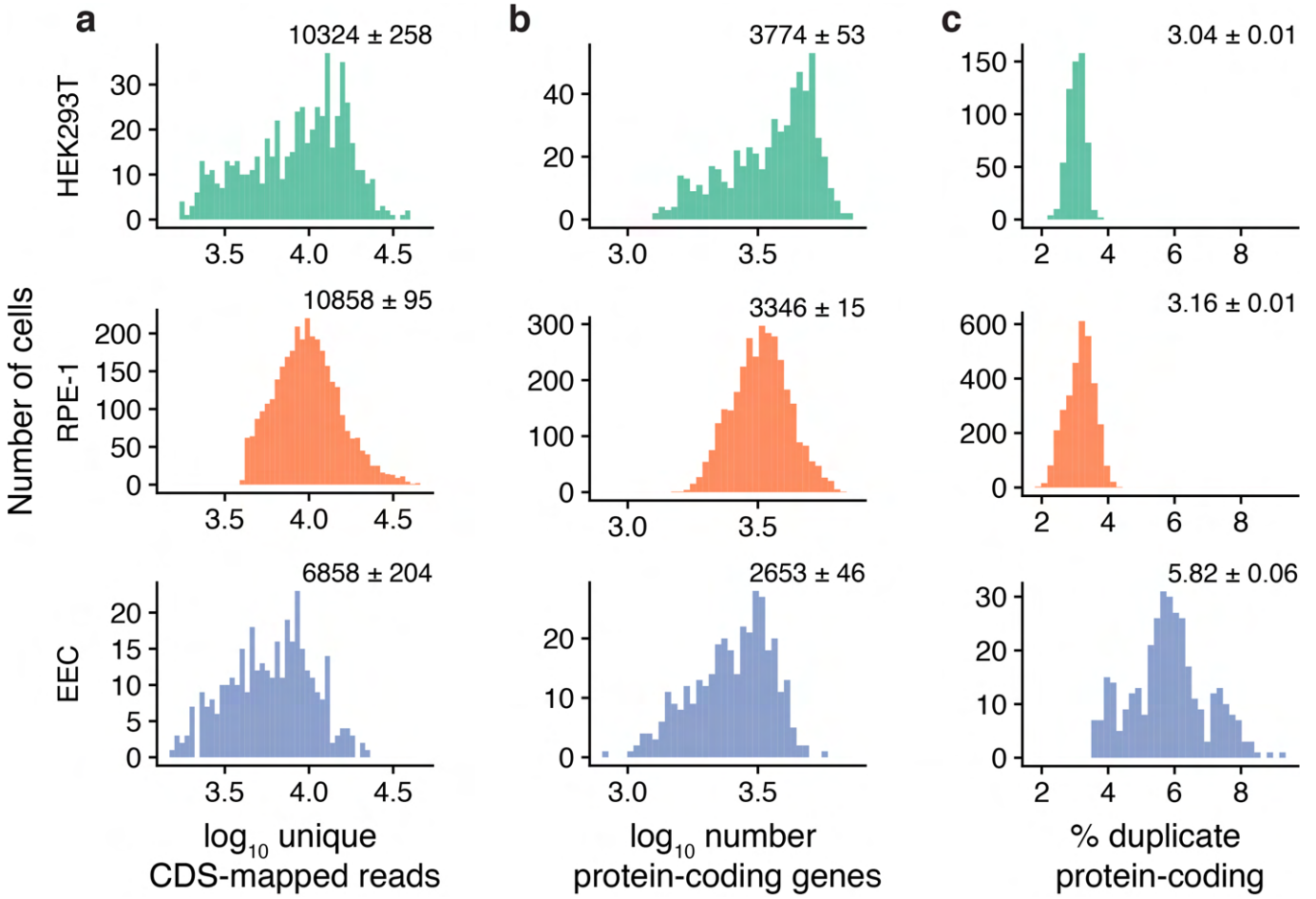
Author contributions M.V. and A.v.O. conceived and designed the project. M.V. developed the experimental protocol and performed single-cell ribosome profiling experiments with help of J.v.d.B. M.V. and A.v.O analysed the data. M.V., J.v.d.B. and A.v.O. discussed and interpreted results. A.A.-R. and H.C. provided research material. M.V. wrote the manuscript with feedback from A.v.O. and J.v.d.B.

Competing interests The technology described here is the subject of a patent application EP20209743 on which M.V. and A.v.O are inventors.

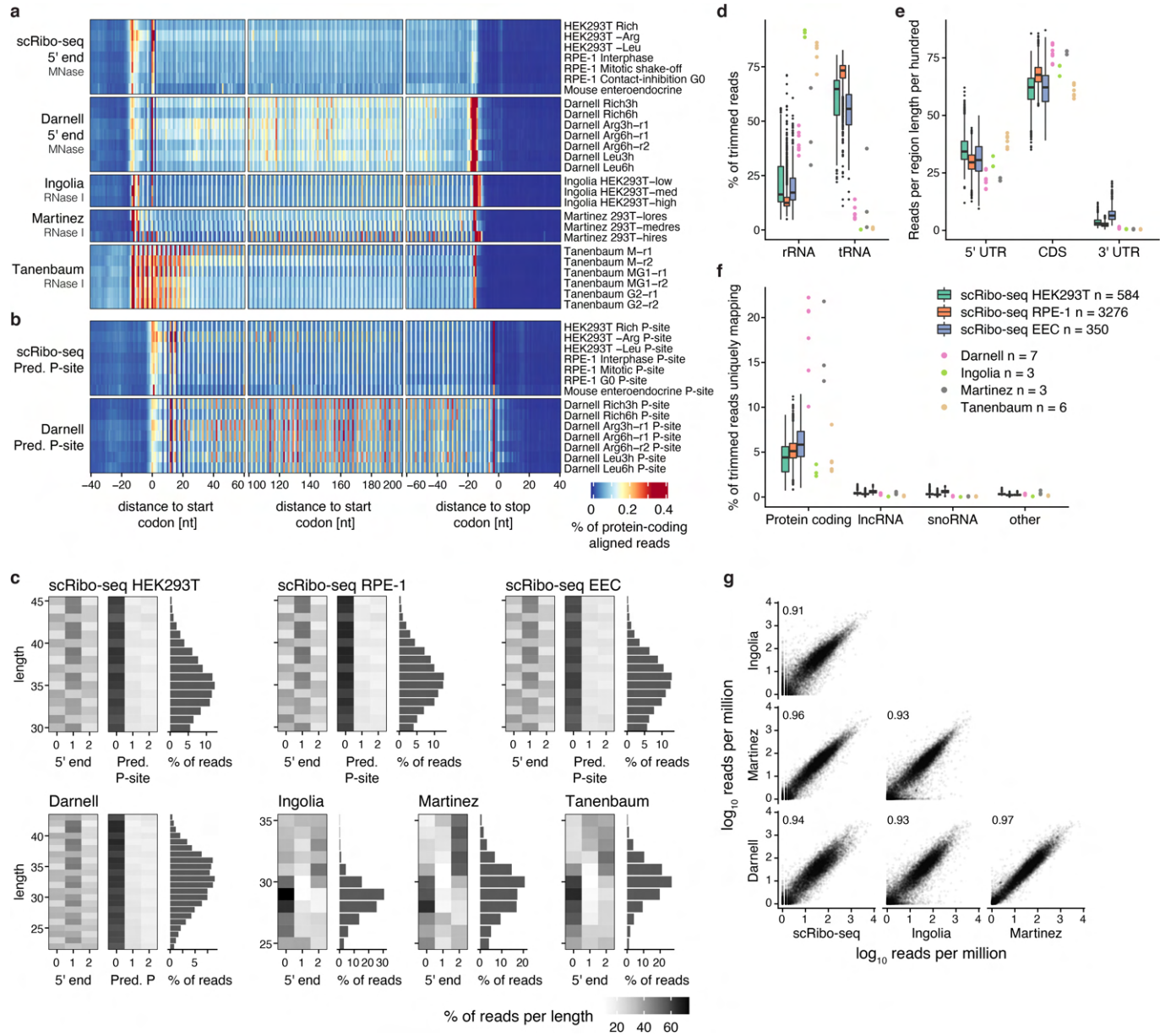
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03887-4.

Correspondence and requests for materials should be addressed to Michael VanInsberghe or Alexander van Oudenaarden.

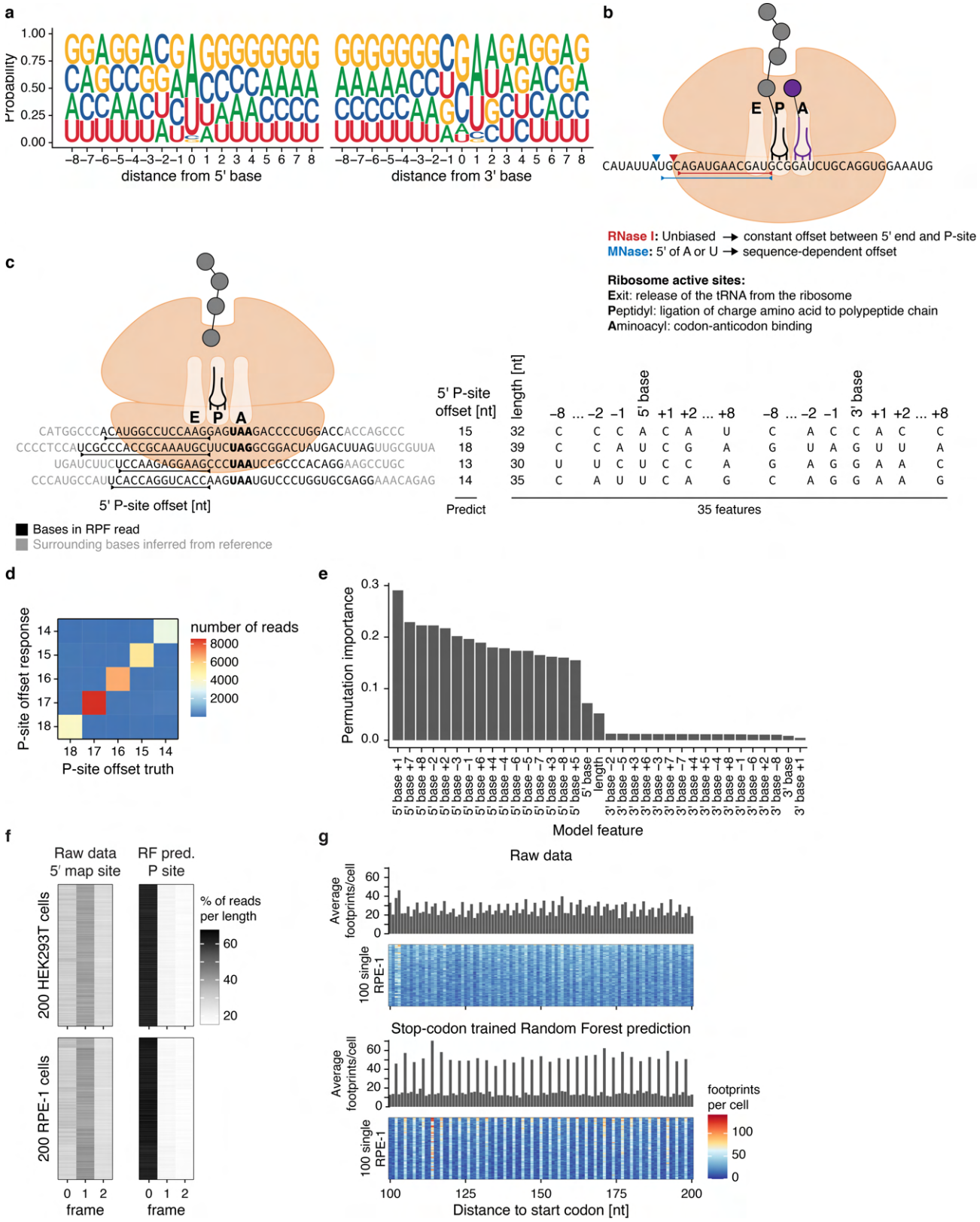
Peer review information *Nature* thanks Arjun Raj, Petra Van Damme and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | Library metrics for scRibo-seq libraries. a, Distributions of the number of unique coding-sequence mapped reads per cell. **b**, Distributions of the number of protein-coding genes detected per cell. **c**, Duplicate rate per cell. The mean ± standard error of each distribution is indicated.

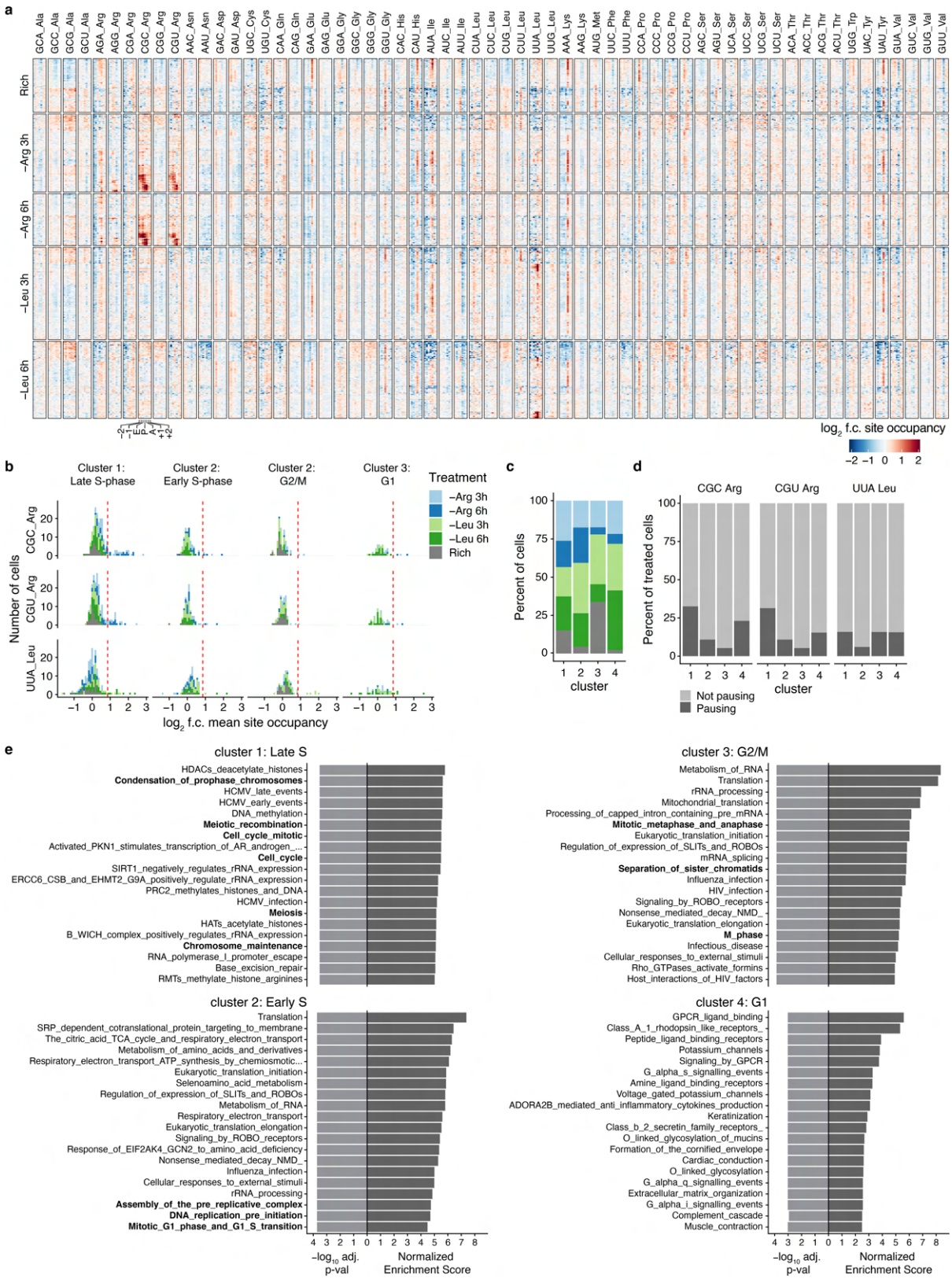


Extended Data Fig. 2 | Comparison of scRibo-seq to conventional ribosomal profiling. a, **b**, Heat maps of the percentage of protein-coding reads per library aligning along metagene regions around the start codon (left), in the CDS (middle), and around the stop codon (right). The mapping coordinate of the 5′ end (**a**), or the random-forest predicted P-site of each read (**b**) is reported. Libraries are from this work (scRibo-seq), and representative bulk ribosomal profiling methods: Darnell[6], using MNase on HEK 293T; Ingolia[8], using RNase I on HEK 293T; Martinez[9], using RNase I on HEK 293T; and Tanenbaum[10], using RNase I on RPE-1. **c**, Frame and read-length distributions of the 5′ end of RPFs and random-forest predicted P-sites averaged across library sets. **d**, Distributions of the percentage of trimmed reads aligning to rRNA and tRNA. **e**, Region-length normalized distributions of RPF mapping frequencies in the 5′UTR, CDS, and 3′UTR regions of protein-coding transcripts. **f**, Distributions of the percentage of trimmed reads that uniquely align to protein coding, lncRNA, snoRNAs, or other biotypes. In the box plots in d-f the middle line indicates the median, the box limits the first and third quartiles, and the whiskers the range. Each point is from a single-cell or bulk library. **g**, Comparisons of the RPF counts per CDS in HEK 293T cells between the different studies. Spearman correlation coefficients for each comparison are indicated.

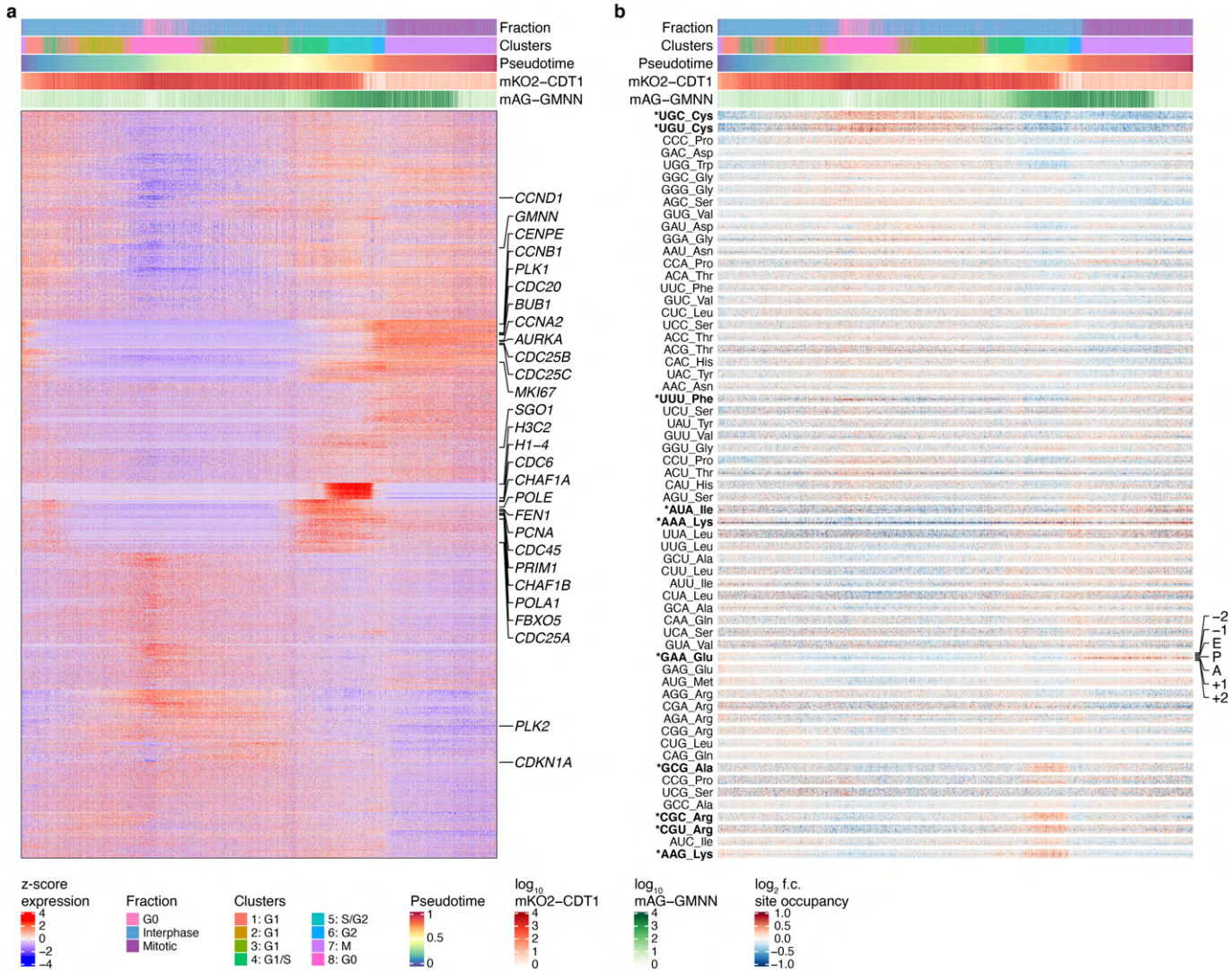


Extended Data Fig. 3 | See next page for caption.

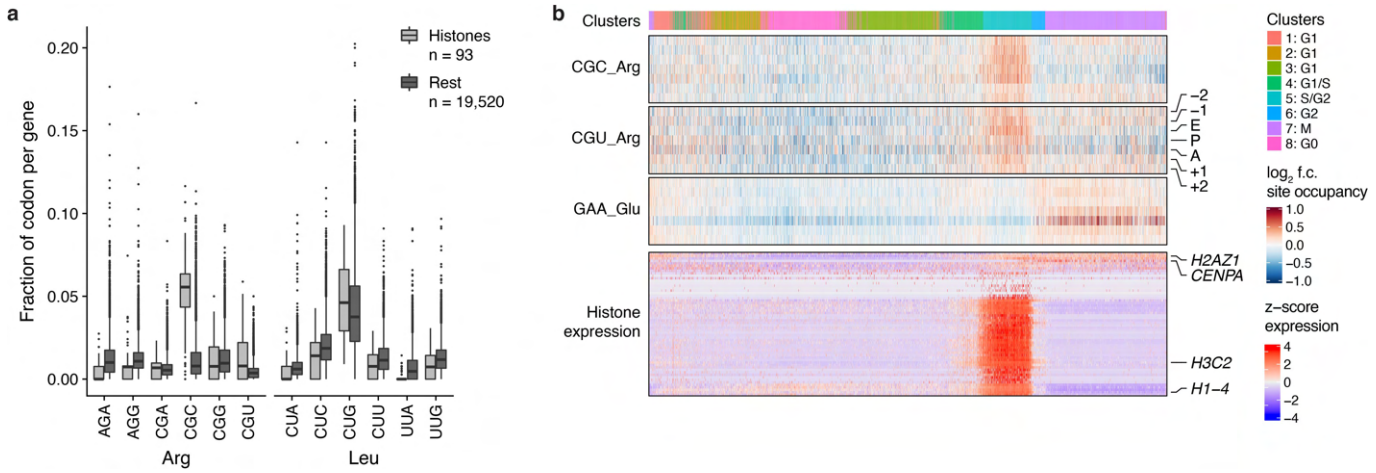
Extended Data Fig. 3 | A Random Forest model corrects the MNase sequence bias to position ribosome active sites within RPF reads. a, Logos of the sequence context around the 5′ and 3′ cut locations. **b**, Schematic illustrating how a nuclease sequence bias can result in a sequence-dependent offset (arrowed lines) between the cut position (triangles) and the ribosome exit, peptidyl, and aminoacyl active sites. Ribosome schematic adapted from ref. [31]. **c**, Schematic describing the parameters used to train the random forest model. Reads spanning a stop codon were used for training. The model predicts the offset between the 5′ end of each read and the P-site based on the read length and the sequence context around each end of the read. **d**, Truth table of the model prediction results on validation data. **e**, Permutation importance of the model features. **f**, Frame distributions of the 5′ end of RPFs and random-forest predicted P-sites in single cells. Both the 5′ and predicted P-sites are uniform between cells and cell types. **g**, Number of footprints per cell along a metagene region within CDS before (top, reads whose 5′ ends align at the given region) and after (bottom, number of predicted P-sites at each location) the random forest correction.



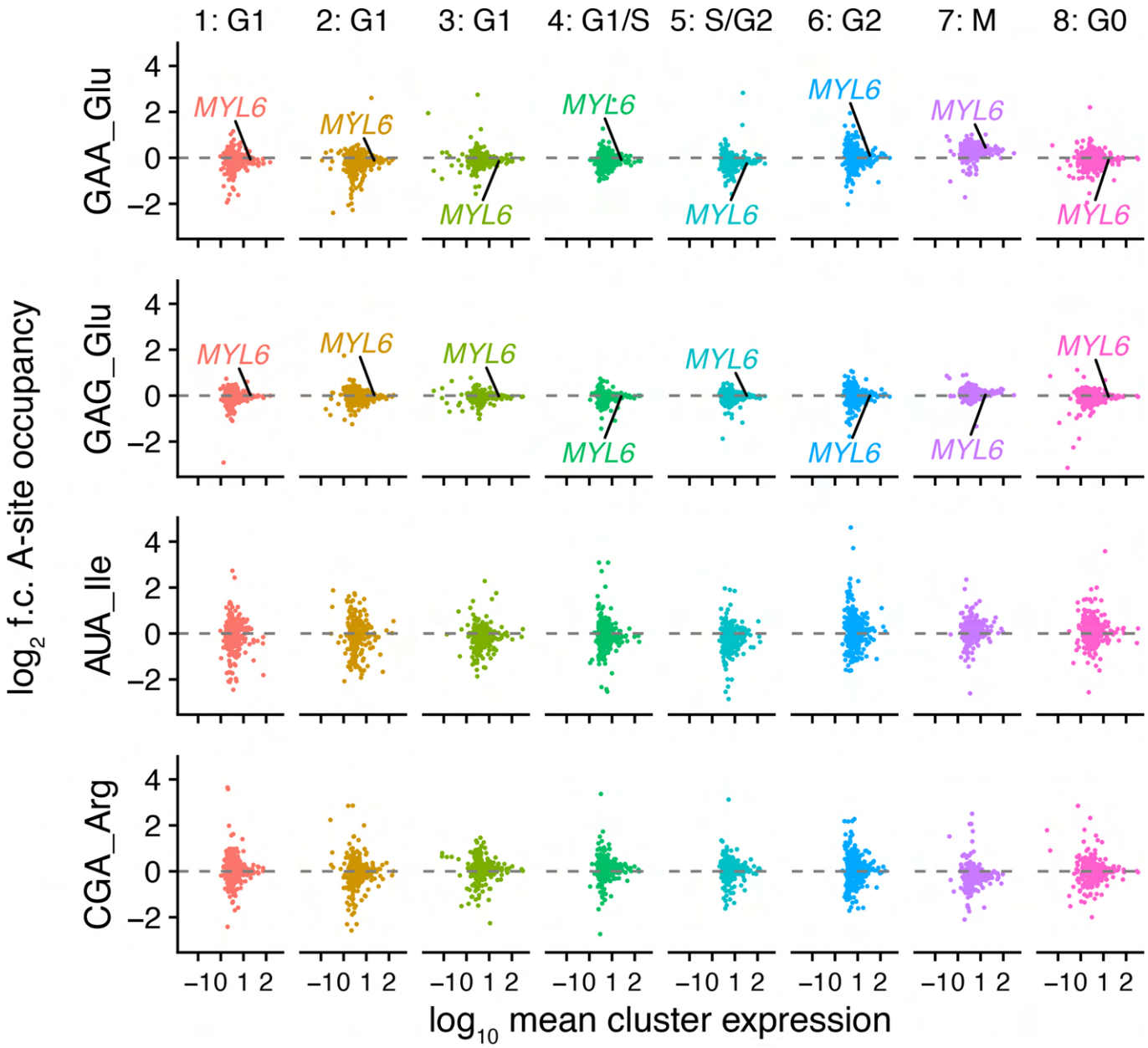
Extended Data Fig. 4 | Ribosome pausing in single cells under amino acid limitation. a, Heat map of the log2 fold change of amino acid occupancy in the RPF active sites. **b**, Distribution of cells exhibiting ribosome pausing in clusters. The threshold used to distinguish pausing cells was calculated as the mean plus 4 standard deviations of the signal of the cells from the rich condition. **c**, Proportions of treatment type per cluster. **d**, Proportions of treated cells that show a pausing response per cluster. **e**, Gene set enrichment analysis[28] on the Reactome Pathway database showing the top twenty categories based on marker genes for HEK 293T cell clusters. Categories associated with the cell cycle are highlighted in bold.



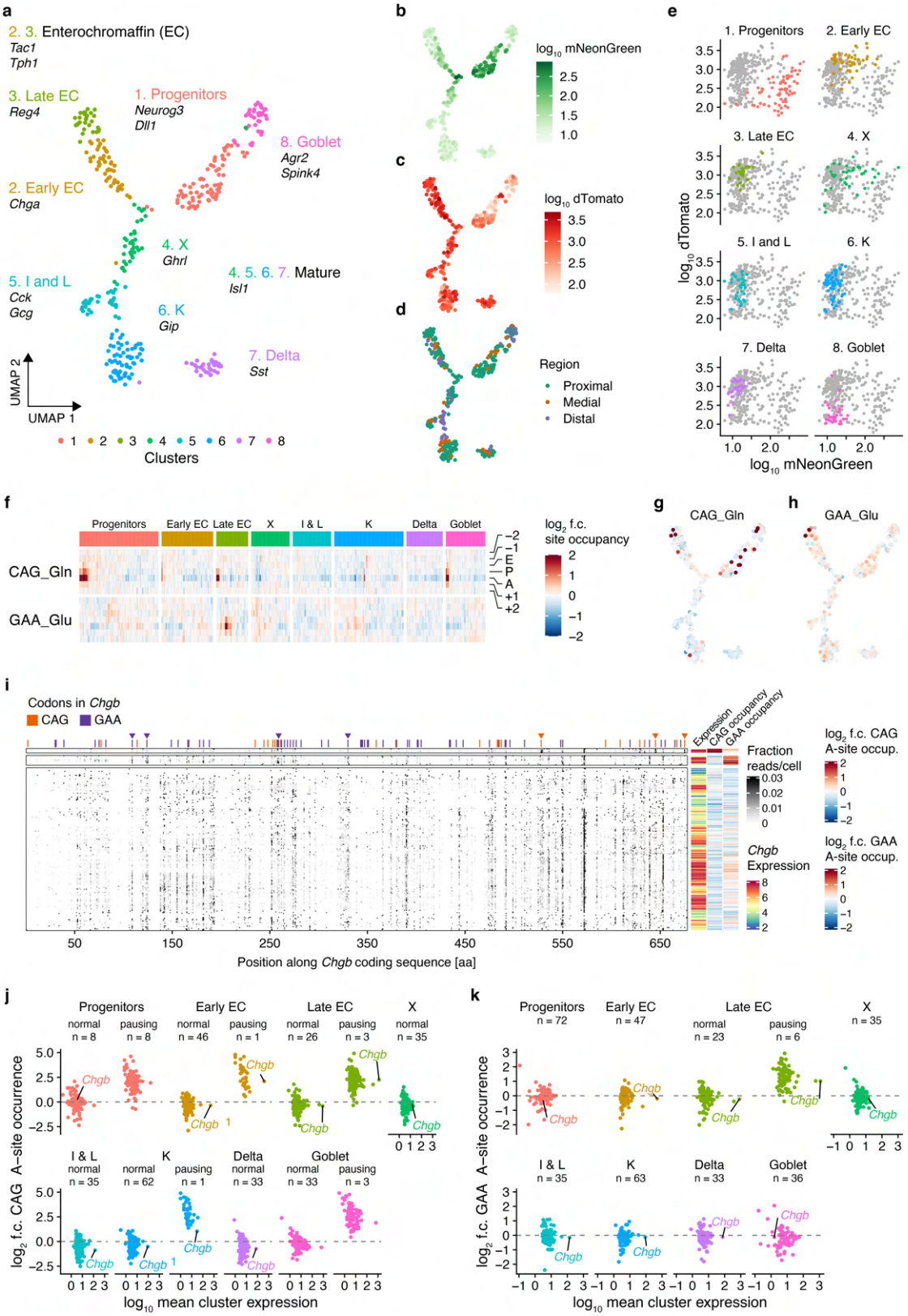
Extended Data Fig. 5 | Marker gene expression and site-specific codon abundance over the cell cycle. a, Heat map of RPF abundance per CDS in hTERT RPE-1 FUCCI cells, showing the translation dynamics of 1,853 significantly differentially translated genes during the cell cycle. Common cell cycle markers are highlighted. **b**, Heat map showing ribosome-site-specific pausing over all codons for hTERT RPE-1 FUCCI cells. Cells are ordered based on cell cycle progression, and codons are clustered based on the average change in the frequency of occurrence across all sites. Codons with significantly different site occupancies between clusters are indicated with an asterisk.



Extended Data Fig. 6 | Ribosome pausing is distinct from changes in codon usage. a, Frequency of arginine and leucine codons in histone genes compared to all other genes. Histone genes (light grey) are highly enriched in CGC and CGU codons compared to other genes. Histone genes were defined as those in HGNC gene group 864. In the box plots the middle line indicates the median, the box limits the first and third quartiles, and the whiskers the range. Each point represents a gene. **b**, Heat map of the fold change in codon occupancy for CGC and CGU codons in the ribosome active sites (top) and the expression of histone genes (bottom) in RPE-1 cells. The site-agnostic increases in CGC and CGU in RPF active sites are synchronous with the increase in translation of histone genes during late S phase (cluster 5, teal). The increases of CGC and CGU codons in all active sites is distinct from the pattern seen in the GAA site occupancies, where the increase is specific to the A site.

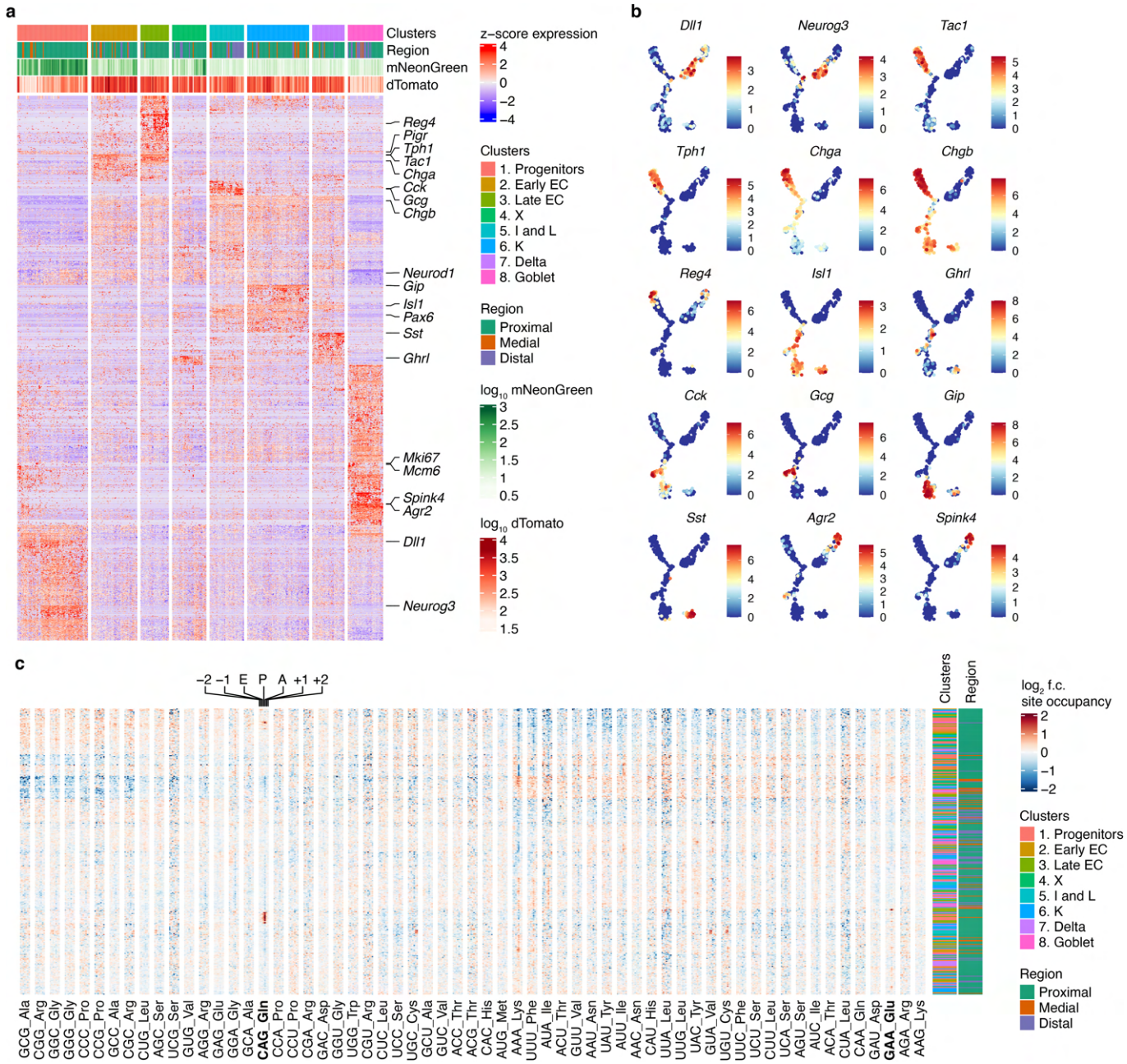


Extended Data Fig. 7 | Scatter plots showing the fold change in gene-wise A-site frequency of occupancy between each cell cluster and the background for the listed codons. The increases (GAA, GAG, and AUA) and decreases (CGA) of the A-site abundance affect the majority of the genes detected across clusters.

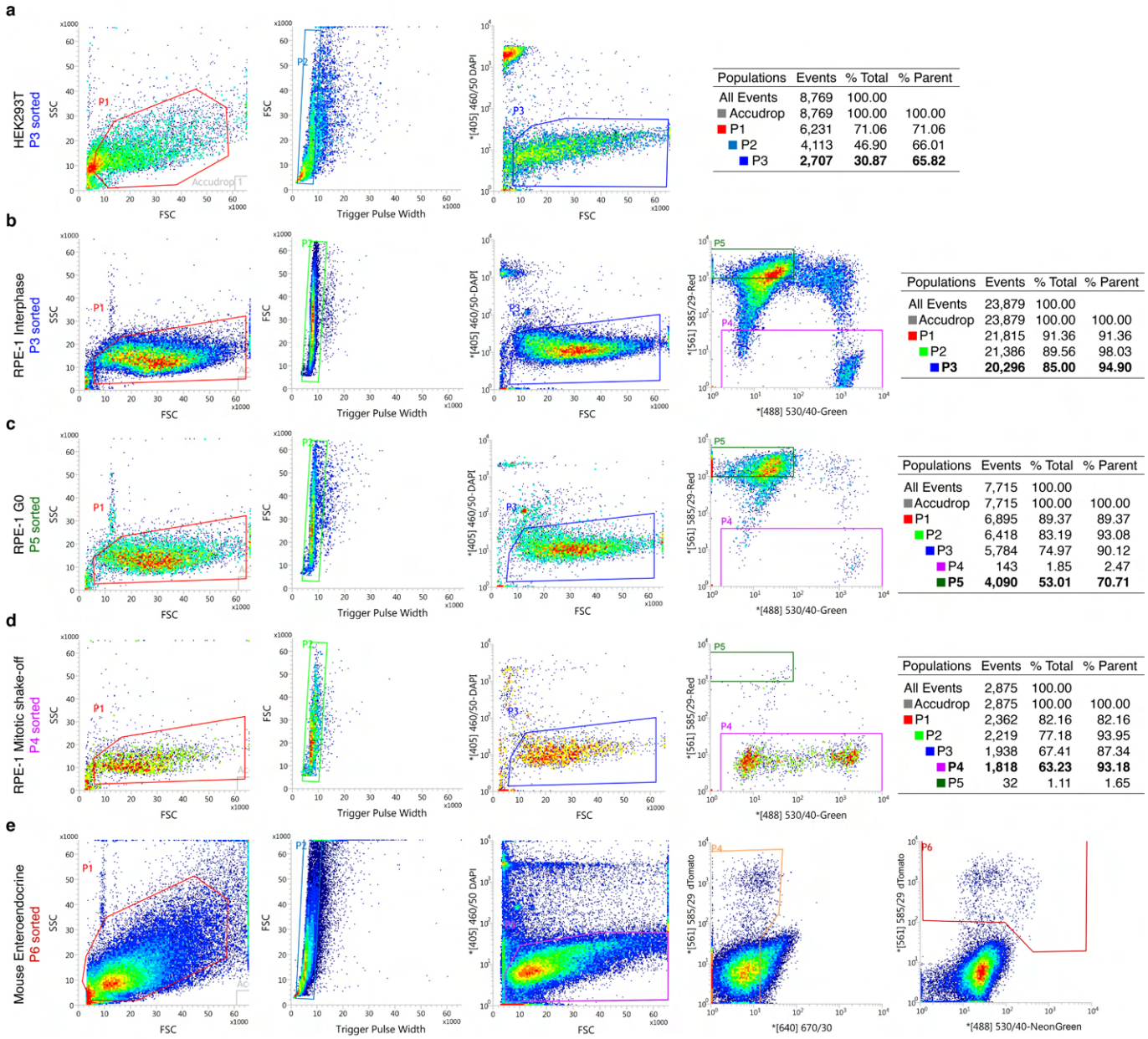


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Single-cell ribosome profiling in primary mouse intestinal EEC cells. a, UMAP (n = 350 cells) generated using the RPF counts per CDS. Corresponding cell types and associated marker genes for each cluster are indicated. **b**, **c**, UMAPs illustrating the fluorescence of the mNeonGreen (**b**) and dTomato (**c**) markers from the bi-fluorescent Neurog3Chrono reporter[24]. **d**, UMAP depicting the intestinal region origin of each cell. As expected, there is no enrichment of the cell types within each region. **e**, Scatter plots of the Neurog3Chrono fluorescence denoting the position of each cell cluster within the FACS space. As expected, progenitor cells show an increased mNeonGreen fluorescence, that changes through a double-positive population to dTomato-positive as EEC cells develop. **f**, Heat map showing ribosome-site-specific pausing over CAG and GAA codons. To remove any effects of the uneven distribution of RPFs along highly translated hormone genes, any gene that was more than an average of 2.5% of the RPFs per cell was removed from this analysis. **g**, **h**, UMAPs showing the CAG (**g**) and GAA (**h**) pausing. **i**, Heat map showing the distribution of RPF A sites along the *Chgb* CDS. Cells are grouped based on their CAG and GAA pausing status. The position of CAG (orange) and GAA (purple) codons within the CDS are denoted as ticks at the top, with shared prominent pausing sites for each codon indicated with inverted triangles. **j**, **k**, Scatter plots showing the fold change in gene-wise A-site frequency of occurrence between the pausing and non-pausing (normal) cells within each cluster.



Extended Data Fig. 9 | Marker genes and codon pausing for EEC cells. a, Heat map of 1,517 genes significantly differentially expressed between the cell clusters. Common EEC marker genes are indicated. **b**, UMAPs (n = 350 cells) showing the expression of common EEC marker and hormone genes. **c**, Heat map showing ribosome-site-specific pausing for all codons in the EEC cells. Cells are clustered based on the profiles across the codons. To remove any effects of the uneven distribution of RPFs along highly translated hormone genes, any gene that was more than an average of 2.5% of the RPFs per cell was removed from this analysis (removed genes: *Chga*, *Chgb*, *Clca1*, *Fcgbp*, *Gcg*, *Ghrl*, *Gip*, *Nts*, *Reg4*, *Sst*).



Populations	Events	% Total	% Parent
All Events	8,769	100.00	
Accudrop	8,769	100.00	100.00
P1	6,231	71.06	71.06
P2	4,113	46.90	66.01
P3	**2,707**	**30.87**	**65.82**

Populations	Events	% Total	% Parent
All Events	23,879	100.00	
Accudrop	23,879	100.00	100.00
P1	21,815	91.36	91.36
P2	21,386	89.56	98.03
P3	**20,296**	**85.00**	**94.90**

Populations	Events	% Total	% Parent
All Events	7,715	100.00	
Accudrop	7,715	100.00	100.00
P1	6,895	89.37	89.37
P2	6,418	83.19	93.08
P3	5,784	74.97	90.12
P4	143	1.85	2.47
P5	**4,090**	**53.01**	**70.71**

Populations	Events	% Total	% Parent
All Events	2,875	100.00	
Accudrop	2,875	100.00	100.00
P1	2,362	82.16	82.16
P2	2,219	77.18	93.95
P3	1,938	67.41	87.34
P4	**1,818**	**63.23**	**93.18**
P5	32	1.11	1.65

Populations	Events	% Total	% Parent
All Events	68 875	100.00	
Accudrop	68 875	100.00	100.00
P1	59 255	86.03	86.03
P2	55 949	81.23	94.42
P3	44 831	65.09	80.13
P4	35 794	51.97	79.84
P6	**864**	**1.25**	**2.41**

Extended Data Fig. 10 | Example gating strategies and population frequencies. a, HEK 293T cells. **b**–**d**, hTERT RPE-1 FUCCI interphase (**b**), contact-inhibition G0 (**c**) and mitotic shake-off fractions (**d**). **e**, Primary mouse EEC cells. Points are pseudocoloured based on density.

nature research

Corresponding author(s): Alexander van Oudenaarden, Michael VanInsberghe

Last updated by author(s): Jun 30, 2021

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing data were collected on the Illumina NextSeq 500 (NextSeq Control Software version 2.2.0.4), using standard software for basecalling (RTA version 2.4.11). Sample demultiplexing was performed using bcl2fastq (v2.20.0.422). FACS data were collected on a BD Influx (BD FACS Sortware version 1.2.0.142).
Data analysis	Data were analyzed using a combination of publicly available and custom software. Publicly available software included: cutadapt (version 3.2), STARSolo (version 2.7.6a), UMI-tools (version 1.1.1), sambamba (version 0.8.0), tRNAscan-SE (version 2.0.7), python (versions 3.7.3 & 3.8.2), GrandPrix (version 0.1), tensorflow (version 2.3.1), gpflow (version 2.1.1.), numpy (version 1.20.0), pandas (version 1.2.2), pysam (version 0.16.0) , R (version 3.6.3), ranger (version 0.12.1), mlr (version 2.17.1), dplyr (version 1.0.2), tidyr (version 1.1.2), ggplot2 (version 3.3.2), cowplot (version 1.1.0), ComplexHeatmap (version 2.5.3), FGSEA (version 1.12.0), Seurat (version 3.2.2), Concorde (version 03.12.19). All custom scripts and conda environment definitions to process raw data and generate figures are available at https://github.com/mvanins/scRiboSeq_manuscript

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

Raw sequencing data, metadata, and count tables have been made available in the Gene Expression Omnibus under the accession number GSE162060.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. The number of cells analyzed was chosen to enable sufficient technical validation and benchmarking of scRibo-seq.
Data exclusions	All raw data are uploaded in public repositories. In downstream analyses, cells were excluded if they did not meet two quality control thresholds: a minimum number of unique reads aligning to protein-coding transcripts, with a minimum fraction of these reads aligning to coding sequences. These thresholds removed cells where the library construction and/or nuclease footprinting failed. The values of these cutoffs were slightly adjusted for each plate and sample type to compensate for differences in sequencing depth and coding-sequence coverage. For single HEK293T cells we required a minimum of 1000 reads mapping to protein-coding sequences and a minimum of 75 % of mapped reads aligning to coding sequences. For RPE-1 cells these thresholds were increased to a minimum of 4500 reads aligning to protein-coding sequences and at at least 85 % of aligning reads stemming from coding regions. For enteroendocrine cells the thresholds were a minimum of 2000 protein-coding reads and 75 % aligning to coding regions. Additionally, primary enteroendocrine cell doublets that were not removed during sorting were filtered using measurements of forward scatter and side scatter collected during index sorting.
Replication	For the HEK293T starvation experiment, data are from one experiment. For the hTERT-RPE-1 cell-cycle experiments, the interphase fraction was measured in three independent experiments, and the G0 and mitotic shake-off in two; all replication attempts were successful. For the primary mouse enteroendocrine cell analyses, data are generated from one experiment using tissue pooled from two independent animals; no obvious differences between animals were observed.
Randomization	No randomization was performed. Batch and plate effects were mitigated by designing sorting plate layouts so that each condition was present on each plate, thereby allowing us to separate biological effects from any potential plate effects.
Blinding	No blinding was performed since we performed unsupervised analysis techniques (e.g., clustering and dimensionality reduction).

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T and hTERT RPE-1 FUCCI cells were both obtained from the Medema lab at the Netherlands Cancer Institute.
Authentication	None of the cell lines were authenticated.
Mycoplasma contamination	All cell lines routinely tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The Neurog3Chrono allele was maintained on a mixed Mus musculus C57BL/6 background. Animals used in the experiments were aged between 8-22 weeks. Both males and females were used for the experiments. Mice were housed in open housing with 14:10h light:dark cycle at 24 °C and 45-70 % relative humidity with food and water ad libitum. The intestines from two adults were pooled together during cell dissociation; randomization and blinding were not performed.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All mouse experiments were conducted under a project license granted by the Dier Experiment Commissie / Animal Experimentation Committee (DEC) or Central Committee Animal Experimentation (CCD) of the Dutch government and approved by the Hubrecht Institute Animal Welfare Body (IvD).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	HEK293T cells: Cells were dissociated to a single-cell suspension by pipetting up and down. The single-cell suspension was washed and resuspended in PBS containing DAPI as a viability stain and BSA to reduce aggregation. Cells were passed through a 20-micron mesh before sorting. hTERT RPE-1 FUCCI cells: Cells were dissociated to a single-cell suspension using TrypLE, washed and resuspended in PBS containing DAPI as a viability stain and BSA to reduce aggregation. Cells were passed through a 20-micron mesh before sorting. Mouse enteroendocrine cells: Mouse enteroendocrine cells were isolated from the intestines of Neurog3Chrono mice, closely following the methods outlined by Gehart et al. Briefly, mouse small intestines were harvested, cleaned, flushed with PBS0, and separated into proximal, medial, and distal sections. Pieces were cut open and villi were scraped off with a glass

	cover slip and discarded. Tissue pieces were then washed in cold PBS0 before transferring to PBS0 with 2 mM EDTA (Gibco), incubated at 4 °C for 30 minutes on a roller, and then vigorously shaken. Detached crypts were pelleted, resuspended in warm TrypLE Select (Gibco), and mechanically disrupted by pipetting to generate single-cell suspensions. Single-cell suspensions were washed 2× in Advanced DMEM/F12 (Gibco), strained with a 20-µm mesh, and resuspended in Advanced DMEM/F12 containing 4 mM EDTA and 1 µg/mL DAPI for sorting.
Instrument	BD Influx
Software	BD FACS Sortware 1.2.0.142
Cell population abundance	Cell population purity and abundance was not explicitly determined after sorting. FACS was primarily used to i) distribute single cells into individual wells of a 384-well plate for subsequent processing, ii) measure the fluorescence of cell-cycle progression markers of these cells, and iii) measure the fluorescence of the Neurog3Chrono markers.
Gating strategy	HEK293T: Doublets, debris, and dead cells were excluded by gating forward and side scatter in combination with the DAPI channel. Viable single cells were sorted. An example gating strategy is provided in Extended Data Figure 13a. hTERT RPE-1 FUCCI: Doublets, debris, and dead cells were excluded by gating forward and side scatter in combination with the DAPI channel. The gating strategy varied between the interphase, G0, and mitotic fractions. Example gating strategies for each are provided in Extended Data Figures 13 b-d. Interphase: viable single cells were sorted. G0: Viable single cells that were mAG negative and mKO2 positive/high were sorted. Mitotic: Viable single cells that were mKO2 negative were sorted. Mouse enteroendocrine: Doublets, debris, and dead cells were excluded by gating forward and side scatter in combination with the DAPI channel. The measurements of dTomato and mNeonGreen were used to select enteroendocrine cells expressing the Neurog3Chrono reporter and DAPI was used to exclude dead cells.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.