TRANSCRIPTION

Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies

Nico Battich*, Joep Beumer, Buys de Barbanson, Lenno Krenning†, Chloé S. Baron‡,
Marvin E. Tanenbaum, Hans Clevers, Alexander van Oudenaarden*

The regulation of messenger RNA levels in mammalian cells can be achieved by the modulation of synthesis and degradation rates. Metabolic RNA-labeling experiments in bulk have quantified these rates using relatively homogeneous cell populations. However, to determine these rates during complex dynamical processes, for instance during cellular differentiation, single-cell resolution is required. Therefore, we developed a method that simultaneously quantifies metabolically labeled and preexisting unlabeled transcripts in thousands of individual cells. We determined synthesis and degradation rates during the cell cycle and during differentiation of intestinal stem cells, revealing major regulatory strategies. These strategies have distinct consequences for controlling the dynamic range and precision of gene expression. These findings advance our understanding of how individual cells in heterogeneous populations shape their gene expression dynamics.

Mammalian cells use diverse strategies to regulate mRNA levels by controlling their synthesis and degradation rates (1, 2). High synthesis or degradation rates allow cells to rapidly respond to extracellular and intracellular signals (2, 3), whereas low degradation rates allow them to integrate transcriptional information over time (4). The extent to which mammalian cells exploit different regulatory strategies during complex dynamical processes such as cell cycle progression or organ formation remains unclear. This is partially due to the difficulty in distinguishing these strategies when only transcript levels are measured.

Additionally, the study of these regulatory strategies in bulk assays is hindered by the presence of heterogeneous cell types in the same tissue and unsynchronized cell states that result from the cell and circadian cycles (1, 5–7). Advances in single-cell RNA sequencing help to resolve cellular heterogeneity (8–14), yet do not provide insights into how the dynamic control of transcription and degradation leads to the observed expression patterns. The kinetic parameters that govern the life of mRNA can be measured by its metabolic labeling during transcription (1, 5, 6). Here, we demonstrate that mRNA labeled with 5-ethynyl-uridine (EU) can be detected in thousands of single cells by sequencing. We determined transcription and degradation rates in heterogeneous and unsynchronized cell popula-

tions and uncovered mRNA control strategies during the cell cycle of human cells and differentiation of mouse intestinal stem cells.

To measure newly synthesized transcripts in single cells, we labeled mRNA by incubating cells with EU, an analog of uridine that can be biotinylated with click chemistry, a method we have named "single-cell EU-labeled RNA sequencing" (scEU-seq) (15). Briefly, after EU incubation, cells were dissociated, fixed, and permeabilized, and EU-labeled RNA was biotinylated in situ. We sorted single cells and generated mRNA/cDNA hybrids using poly-T primers containing a cell barcode, a unique molecular identifier (UMI), a 5′ sequencing adapter, and the T7 promoter. Cells were pooled, EU labeled and unlabeled hybrids were separated using streptavidin magnetic beads, and libraries were generated for both fractions (Fig. 1A). The UMI counts for labeled mRNA were higher in EU-treated cells compared with dimethyl sulfoxide (DMSO)–treated cells or empty control wells, resulting in a high signal-to-noise ratio and low across-well cross-contamination rates (Fig. 1B). Only 12 of 11,848 detected genes were affected by the EU treatment itself (fig. S1A). When we compared the total mRNA (unlabeled and EU-labeled UMIs; total UMIs) from EU-treated versus DMSO-treated RPE1-FUCCI cells, we found high recovery efficiency (99.5 ± 0.4%) of labeled mRNA (fig. S1B). After 120 min of EU incubation, the labeled mRNA fraction was on average 8.9 ± 0.7% (fig. S1B), which agrees with an expected average production of 8 to 10% of the transcriptome during a period of 2 hours in unsynchronized cells with a cell cycle length of ~20 to 24 hours (16).

To assess whether scEU-seq specifically enriches transcripts synthesized during the EU-labeling window, we performed pulse and chase experiments varying either the EU incubation time or the length of a chase phase with uridine

(U) after EU treatment for 22 hours (Fig. 1C and fig. S1, C to G). As expected, we detected an increase in labeled UMIs as a function of the EU pulse length (Fig. 1D) and a decrease as a function of the U chase length (Fig. 1E). We could still detect significantly higher UMI counts for very short labeling times (15 and 30 min) compared with the DMSO control (Fig. 1D and fig. S1G). In these short EU pulses, we found that labeled UMIs were enriched in transcripts that contained unspliced introns (Fig. 1F).

Next, we incubated K562 cells at either 37°C or 42°C for a period of 45 min in the presence of EU or DMSO. The differential gene expression signature upon heat shock was more pronounced in the fraction of EU-labeled mRNAs compared with the unlabeled fraction or with cells treated with DMSO (fig. S2). Consistently, the functional annotation analysis (17) for up-regulated genes in the EU-labeled fraction revealed an enrichment for genes encoding heat shock or stress response proteins (fig. S2B). In addition, UMIs of these stress response genes represented a large percentage in the EU-labeled samples but not in the DMSO-treated or unlabeled controls.

Using data from the scEU-seq pulse and chase experiments, we can estimate the synthesis rate κ and the degradation rate constant γ for all detected transcripts. Furthermore, we can place individual cells along cell cycle or differentiation trajectories and thus infer how synthesis and degradation rates change over time. We first estimated κ and γ with high resolution along the mammalian cell cycle. For each of the 5422 cells that passed quality controls in the pulse and chase experiments, we calculated the relative position along the cell cycle using the Geminin-GFP and the Cdt1-RFP signals from the FUCCI system (fig. S3, A and B) (18). The expression of known cell cycle markers followed the expected pattern relative to the Geminin-GFP and Cdt1-RFP (18), whereas the housekeeping gene HPRT1 displayed constant expression during the cell cycle (Fig. 2A and fig. S3C). The level of labeled transcripts of cell cycle–controlled genes changed as a function of the cell cycle (fig. S4 and fig. S5), with the total UMI counts per cell approximately doubling during one cell cycle (fig. S6, A and B). These results suggest that our estimation of the cell cycle progression in single cells is accurate.

To fit κ and γ to the experimental datasets, we simulated the dynamics of the pulse and chase experiments and quantified the accuracy of the fitting procedures. The simulations defined the range of κ and γ values for which we can accurately determine these rates, and demonstrated that a model that does not assume steady-state dynamics of gene expression is more fitting for our datasets (fig. S7). Next, we used the cell cycle progression estimates to pool cells from different EU-labeling

Oncode Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, Netherlands.
*Corresponding author. Email: n.battich@hubrecht.eu (N.B.); a.vanoudenaarden@hubrecht.eu (A.v.O.)
†Present address: Division of Cell Biology, Oncode Institute, Netherlands Cancer Institute, Amsterdam, the Netherlands.
‡Present address: Stem Cell Program and Division of Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Howard Hughes Medical Institute, Harvard Medical School, Harvard Stem Cell Institute, Stem Cell and Regenerative Biology Department, Harvard University, 1 Blackfan Circle, Boston, 02115 MA, USA.

time points and fitted κ and γ using the non–steady-state model (Fig. 2B and fig. S8) on 528 genes that showed high expression changes during the cell cycle (fig. S9). Separate fits of the pulse and chase experiments with the non–steady-state model resulted in accurate γ values, but simultaneous fits of both experiments further reduced uncertainty (Fig. 2B and fig. S10, A and B). We averaged the degradation rates over the cell cycle and obtained good agreement between our data and a published dataset (Spearman r = 0.585, fig. S10C) (*5*). The estimated values of κ and γ for the 528 genes allowed us to predict their transcript levels at any position during cell cycle progression. Our predictions matched measurements of CEL-seq2 (*19*), an independent single-cell mRNA-sequencing method, for expression changes along the cell cycle (median correlation of 0.730 for 528 genes; Fig. 2C and fig. S10D).

We found widespread changes of both synthesis and degradation rates during the cell cycle (Fig. 2, B and C). Clustering of the expression levels, synthesis rates, and degradation rate constants with self-organizing maps (SOMs) revealed distinct strategies of mRNA regulation during the cell cycle (Fig. 2, B and D). To find common properties between the different strategies regardless of the position of the expression peak, we computed the cosine similarity between synthesis and degradation dynamics (Fig. 2B and fig. S11A). We observed three types of regulatory strategies during the cell cycle: cooperative, neutral, and destabilizing. The cooperative strategy describes an increase in the synthesis rate that is accompanied by a decrease in the degradation rate constant and vice versa, thus having a negative cosine similarity. The neutral strategy is characterized by small relative changes in the degradation rate constant compared with the synthesis rate. The destabilizing strategy is characterized by a simultaneous increase or decrease of the synthesis rate and the degradation rate constant, resulting in positive cosine similarity. Among the groups of genes that follow the cooperative strategy, we found a subset of genes that have an expression peak in G_2 and are involved in microtubule spindle assembly and mitosis regulation (strategy group B); genes with a functional enrichment for signaling and protein phosphorylation (group F); and genes that are expressed during S phase and are involved in DNA replication, repair, and maintenance (group D; Fig. 2C and fig. S11B). We validated the changes of the degradation rate constant during the cell cycle for these genes (groups B, D, and F) by performing a bulk chase experiment of 500 pooled cells gated for G_1, S, or G_2 (Pearson r between 0.290 and 0.611; fig. S11C). Genes that follow the neutral strategy were functionally enriched in microtubule activity (group A), homology recombination repair (group E), or cytokine activity, G_1/S transition, and DNA replication initiation (group C; Fig. 2C and fig. S11B). Further simulations in which we varied the regimes of κ and γ throughout the cell cycle (fig. S12, A to C) validated that we can accurately determine the strategy type for most of the parameter combinations tested (fig. S12, D to G). These results indicate that genes with similar cellular functions tend to be controlled by similar strategies and may share posttranscriptional regulators.

We next investigated the change in predicted expression by assuming either a model with constant γ or constant κ (Fig. 2, D and E). These constants were chosen to match the expression averaged over the cell cycle as observed in the experimental data. The model assuming dynamic γ and κ could accurately
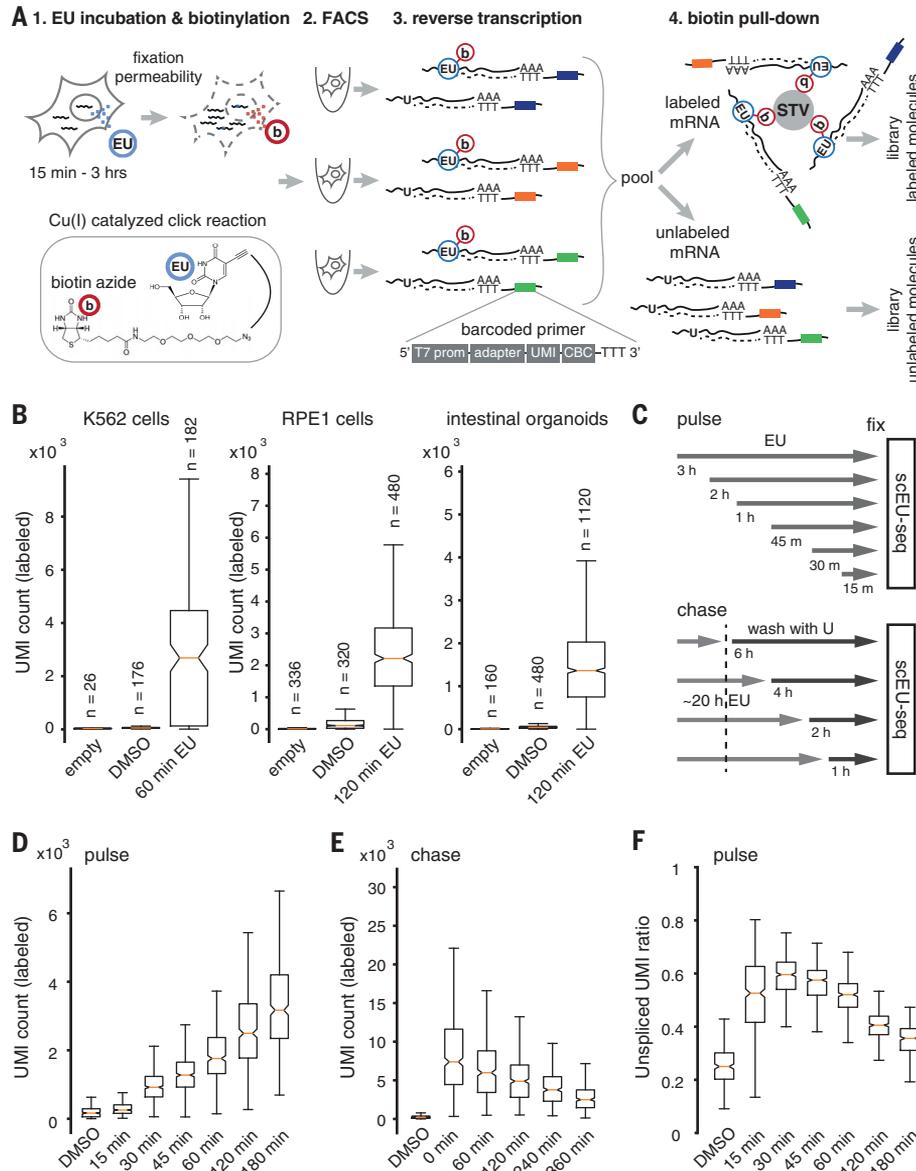
Fig. 1. Single-cell EU RNA sequencing (scEU-seq). (**A**) Schematics of the scEU-seq workflow. (**B**) Boxplots showing UMI counts of labeled mRNAs per well either containing single cells or left empty. Cell types are indicated. No cutoff was applied to the UMI counts per well. Signal-to-noise ratios are 49.74 (P = 2.4 × 10^{-26}) for K562 cells, 19.69 (P = 2.4 × 10^{-77}) for FUCCI-expressing RPE1 cells, and 32.48 (P = 1.7 × 10^{-157}) for cells derived from mouse intestinal organoids; P values are from a Mann–Whitney U test. Estimated cross-contamination rates per well are 37.31 ± 4.11, 21.13 ± 1.80, and 8.26 ± 0.87 (mean UMIs per cell ± SEM) for K562, RPE1-FUCCI, and organoid cells, respectively. (**C**) Design of pulse and chase experiments. (**D**) Boxplots showing UMI counts of labeled mRNAs per cell. RPE1-FUCCI cells were treated with EU for the indicated times. (**E**) As in (D), but RPE1-FUCCI cells were treated with EU for 22 hours and then washed and treated with U for the indicated times. Cells shown in (D) and (E) were filtered as described in (*15*). (**F**) Fractions of UMI of labeled mRNAs containing unspliced introns for RPE1-FUCCI cells.

predict the dynamic range and the expression timing of our CEL-seq2 dataset. However, when either rate was assumed constant, we observed changes in both expression properties. As expected, the synthesis rate κ had an impact on the dynamic range of all genes (Fig. 2F), whereas the impact on timing was strongest for the strategy groups A and G, which show constant degradation rates during the cell cycle (Fig. 2G). Strong effects on the dynamic range could be observed by assuming a constant γ for the groups of genes showing cooperative strategies (groups B, D, and F; Fig.

2F). In addition, for three of the four clusters in group B, the effects on expression timing obtained by assuming constant degradation were similar to the results obtained by the constant synthesis model (Fig. 2G). These results imply that the degradation and synthesis rates are coordinated to achieve precise expression dynamics during the cell cycle.

Next, we asked whether scEU-seq could reveal similar mRNA regulatory strategies during cellular differentiation. We used intestinal organoids expressing the GFP-Lgr5 reporter in intestinal stem cells (20, 21) (Fig. 3A) and

performed a pulse experiment using an EU incubation time of 120 min and chase experiments with a 0-, 45-, or 360-min U chase phase. The UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) representation of the 3831 cells that passed quality controls places stem cells in the center and shows two branches representing the differentiation trajectories of enterocytes and secretory cells (Fig. 3, B to D). We did not observe batch effects between the experiments (fig. S13A). The expression of the GFP-Lgr5 fusion closely matches the measured expression of the *Lgr5*

Fig. 2. scEU-seq reveals mRNA control strategies during the cell cycle. (**A**) Scatter plot of the Geminin-GFP- and Cdt1-RFP–corrected signals of RPE1-FUCCI cells (n = 5422 cells). Expression levels (total UMI counts per cell) of four example genes are indicated in blue. (**B**) Clustered heat maps of estimated synthesis and degradation rates. Leftmost panel shows the cosine similarity (s) between the rates (n = 528 genes). (**C**) As in (B) but showing the observed expression levels (left, data generated using CELSeq2) and predicted expression levels (right) along cell cycle progression. (**D**) Schematics of the calculation of the dynamic range (top) and the timing of the expression peak (bottom). (**E**) Density plot of the peak timing distance against the dynamic range of the predicted relative to the observed expression for models with dynamic synthesis and degradation rates (black, left), a constant synthesis rate (blue, middle), and a constant degradation rate (red, right). Top panels compare the distributions of peak timing distances (blue versus black: $P = 1.05 \times 10^{-22}$, red versus black: $P = 6.67 \times 10^{-8}$, F test for variance). Rightmost panel compares the distributions of dynamic ranges (blue versus black: $P = 9.98 \times 10^{-85}$, red versus black: $P = 1.63 \times 10^{-58}$, Wilcoxon test, n = 528 genes). (**F**) Median absolute differences (delta) in dynamic range between the constant synthesis model [blue in (E)] and the full dynamic model [black in (E)] and between the constant degradation model [red in (E)] and the full dynamic model [black in (E)]. (**G**) As in (F) but for the peak timing distance.
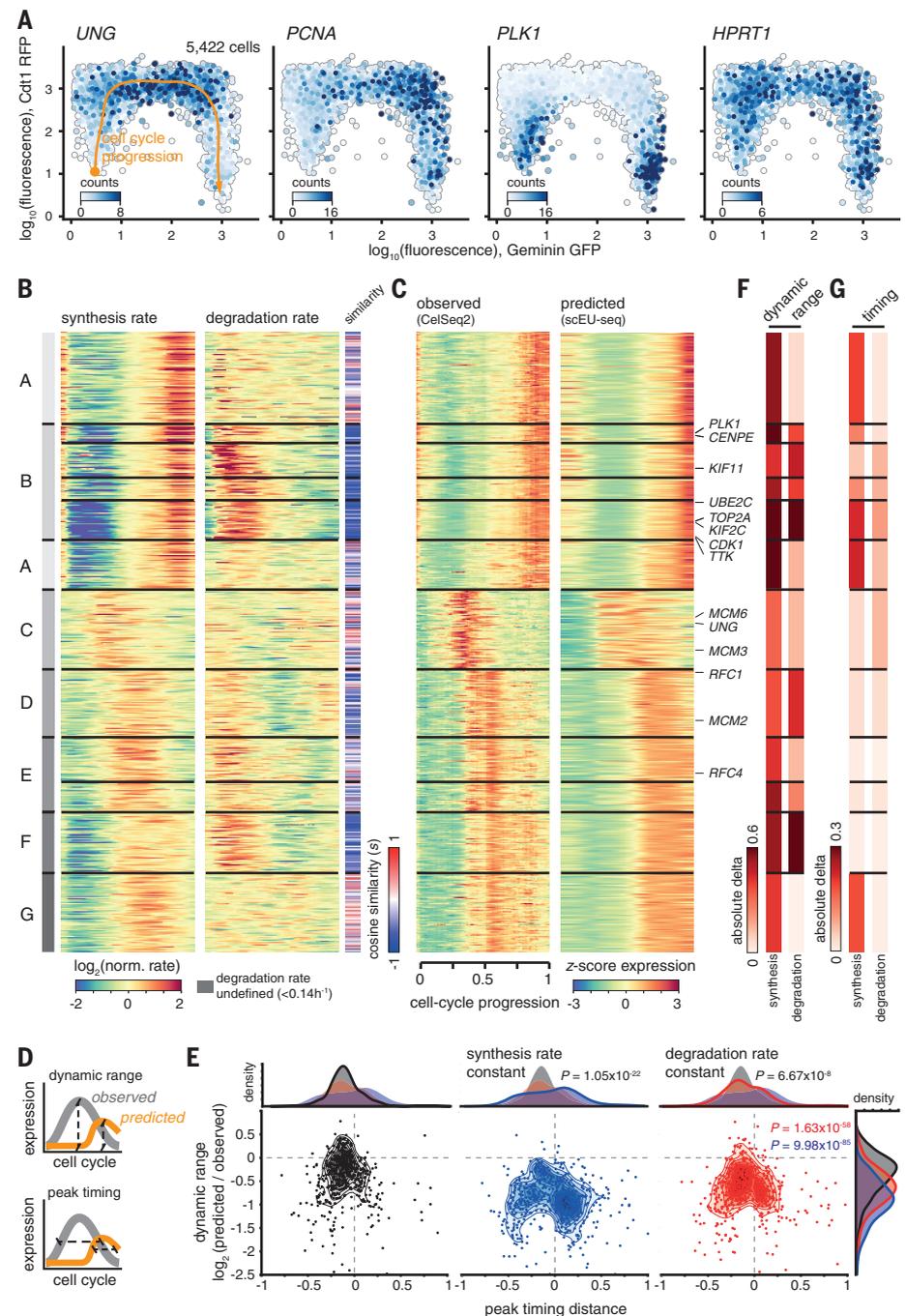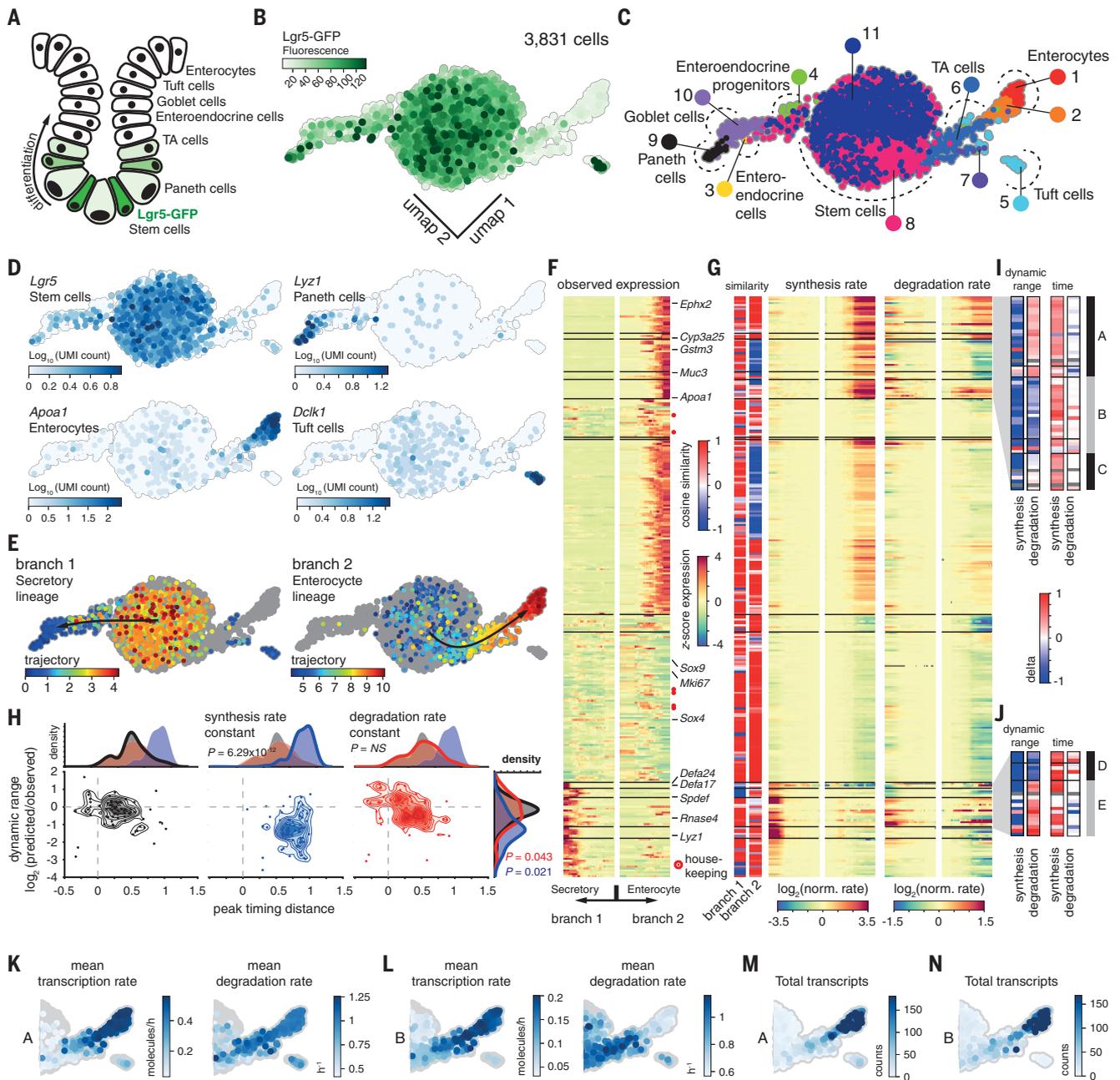
Fig. 3. mRNA control strategies during intestinal organoid differentiation.
(**A**) Schematics of intestinal organoid crypts. (**B**) UMAP showing the expression levels of Lgr5-GFP (green) in 3831 cells from intestinal organoids. (**C**) UMAP showing clusters of cells with similar gene expression (SOM analysis, cluster number indicated) and their respective cell identity. (**D**) UMAP showing the expression levels (blue) of four genes that are markers for stem cells, Paneth cells, enterocytes, and tuft cells, respectively. (**E**) UMAP showing the Monocle2 differentiation branches 1 and 2, as indicated by arrows. Colors indicate the monocle trajectory values. (**F**) Heat maps of the observed expression levels along the two differentiation branches, the secretory lineage into Paneth cells (branch 1) and the enterocyte lineage (branch 2) (n = 301 genes). Red dots mark the position of housekeeping genes. (**G**) As in (F) but showing the estimated synthesis (left panels) and degradation rates (right panels). Genes are clustered and the cosine similarity is indicated independently for the two branches. Strategies with strong changes in the synthesis and the degradation rates are highlighted in (I) and (J). (**H**) Density plot of the peak timing distance against the dynamic range of the

predicted relative to the observed expression for models with dynamic synthesis and degradation rates (black, left), constant synthesis rate (blue, middle), and constant degradation rate (red, right) [n = 72 genes with strong cooperative and destabilizing strategies (groups A, B C, D and E); see (I) and (J)]. Top panels compare the distributions of peak timing distances (blue versus black: $P = 6.29 \times 10^{-12}$, red versus black: P = nonsignificant, Wilcoxon test). Rightmost panel compares the distributions of dynamic ranges (blue versus black: P = 0.021, red versus black: P = 0.043, F test for variance, n = 72 genes). (**I** and **J**) Delta values of data shown in (H) for genes with strong cooperative and destabilizing strategies (groups A, B C, D, and E). Shown are differences in dynamic range and timing, respectively, between the constant synthesis model [blue in (H)] and the full dynamic model [black in (H)] and between the constant degradation model [red in (H)] and the full dynamic model [black in (H)]. (**K**) Mean synthesis and degradations rates for genes in group A. (**L**) Mean synthesis and degradation rates for genes in group B. (**M**) Total number of UMIs detected for genes in group A. (**N**) Total number of UMIs detected for genes in group B.

transcript (Fig. 3, B and D). We clustered single-cell expression levels using SOMs and identified stem cells (clusters 8 and 11), fully differentiated cells (clusters 1, 3, 5, 9, and 10), and potentially intermediate stages (clusters 10, 4, 6, and 2; Fig. 3C).

We selected 295 genes that showed a high coefficient of variation compared with the mean expression and differences in expression between the stem cells and cells in intermediate or differentiated stages (fig. S13, B and C). We used monocle2 (22) to sort cells along the differentiation trajectories of secretory cells (branch 1) and enterocytes (branch 2; Fig. 3E and fig. S14, A to C). We added six housekeeping genes as a control and calculated the synthesis rates and degradation rate constants throughout differentiation for these 301 genes (fig. S14, D to F).

We used SOMs to cluster genes by their expression level, synthesis rates, and degradation rate constants. The results demonstrate that cells use both the synthesis and degradation rate to control gene expression during differentiation. The housekeeping genes clustered separately from differentially regulated genes between branches 1 and 2 (Fig. 3, F and G, and fig. S15A). For 24% of genes (72 genes), the degradation rate changed during differentiation, whereas the synthesis rate increased, displaying cooperative and destabilizing strategies, respectively, as observed for the cell cycle. Among the genes with destabilizing strategies (group A), we identified functional enrichment for oxidoreductase activity and drug metabolism (fig. S15B). These genes were up-regulated in enterocytes and six of them belong to the cytochrome P450 family localized to the endoplasmic reticulum (23). Gene groups with cooperative strategies (group D and B) were enriched for genes encoding secreted proteins or components of the endoplasmic reticulum or Golgi complex (fig. S15B).

When we analyzed the gene expression dynamics as before, assuming a constant synthesis rate, we observed that the dynamic range decreased and the expression timing and dynamics changed (Fig. 3H). By contrast, a constant degradation rate had little effect on the timing of the expression peak but increased the variance of the dynamic range; although the dynamic range of genes with destabilizing strategies increased (groups A and E), it decreased for genes with cooperating strategies (groups B and D; Fig. 3, I and J). In agreement with this, we found that the absolute increase in the synthesis rate of genes in group A was higher than those in group B (Fig. 3, K and L), whereas the expression levels in both groups changed along the differentiation trajectory with similar dynamics and magnitude (Fig. 3, M and N). This effect is explained by the stabilization of transcripts in group B toward the end of differentiation branch 2 (Fig. 3L and fig. S15C).

Here, we show that cells use cooperative, neutral, or destabilizing strategies to actively regulate gene expression during the cell cycle and during differentiation. Both synthesis and degradation rates control the accuracy and precision of the dynamic range and the timing of the expression peak. By contrast, during differentiation, the degradation rate seems to affect only the dynamic range of expression, whereas the timing is fully encoded by the dynamics of the mRNA synthesis rate. Thus, our data support findings that the modulation of mRNA degradation rates plays a role in mammalian cellular homeostasis such as T cell homeostasis (24) and differentiation of mammalian embryonic stem cells (25).

REFERENCES AND NOTES

1. B. Schwalb et al., Science 352, 1225–1228 (2016).
2. M. Rabani et al., Cell 159, 1698–1710 (2014).
3. O. Shalem et al., Mol. Syst. Biol. 4, 223 (2008).
4. S. C. Little, M. Tikhonov, T. Gregor, Cell 154, 789–800 (2013).
5. H. Tani et al., Genome Res. 22, 947–956 (2012).
6. M. Rabani et al., Nat. Biotechnol. 29, 436–442 (2011).
7. A. Raghavan et al., Nucleic Acids Res. 30, 5529–5538 (2002).
8. T. Hashimshony, F. Wagner, N. Sher, I. Yanai, Cell Rep. 2, 666–673 (2012).
9. D. A. Jaitin et al., Science 343, 776–779 (2014).
10. A. B. Rosenberg et al., Science 360, 176–182 (2018).
11. D. Grün et al., Nature 525, 251–255 (2015).
12. E. Z. Macosko et al., Cell 161, 1202–1214 (2015).
13. A. M. Klein et al., Cell 161, 1187–1201 (2015).
14. B. Pijuan-Sala et al., Nature 566, 490–495 (2019).
15. See supplementary materials.
16. T. Zerjatke et al., Cell Rep. 19, 1953–1966 (2017).
17. W. da Huang, B. T. Sherman, R. A. Lempicki; W. Huang da, Nat. Protoc. 4, 44–57 (2009).
18. L. Krenning, F. M. Feringa, I. A. Shaltiel, J. van den Berg, R. H. Medema, Mol. Cell 55, 59–72 (2014).
19. T. Hashimshony et al., Genome Biol. 17, 77 (2016).
20. T. Sato et al., Nature 459, 262–265 (2009).
21. H. Tian et al., Nature 478, 255–259 (2011).
22. X. Qiu et al., Nat. Methods 14, 979–982 (2017).
23. F. Xie, X. Ding, Q. Y. Zhang, Acta Pharm. Sin. B 6, 374–383 (2016).
24. S. Geula et al., Science 347, 1002–1006 (2015).
25. P. J. Batista et al., Cell Stem Cell 15, 707–719 (2014).

Science

Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies

Nico Battich, Joep Beumer, Buys de Barbanson, Lenno Krenning, Chloé S. Baron, Marvin E. Tanenbaum, Hans Clevers and Alexander van Oudenaarden

RNA life span at single-cell resolution

RNA transcripts are an easily accessed representation of gene expression, but we lack a comprehensive view of the life span of RNA within the single cell. Battich *et al.* developed a method to sequence messenger RNA labeled with 5-ethynyl-uridine (EU) in single cells (scEU-seq), which allows estimation of RNA transcription and degradation rates. When examining intestinal organoid cells, scEU-seq data can be used to discern between transcription and degradation during development, indicating that this method can be applied to better understand the relationship between gene expression and RNA degradation during development.

Science, this issue p. 1151

ARTICLE TOOLS	http://science.sciencemag.org/content/367/6482/1151
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2020/03/04/367.6482.1151.DC1
REFERENCES	This article cites 38 articles, 7 of which you can access for free http://science.sciencemag.org/content/367/6482/1151#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service