Molecular Cell
Spotlight

RNA Velocity: Molecular Kinetics from Single-Cell RNA-Seq

Valentine Svensson[1,*] and Lior Pachter[1,*]
[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
*Correspondence: vale@caltech.edu (V.S.), lpachter@caltech.edu (L.P.)
https://doi.org/10.1016/j.molcel.2018.09.026

Applying a kinetic model of RNA transcription and splicing, La Manno et al. (2018) predict changes in mRNA levels of individual cells from single-cell RNA-seq data.

As the central dogma of molecular biology came into view in the 1950s, researchers started modeling the kinetic processes of biosynthesis using differential equations for reactions. In Gorini and Maas (1957), a first-order differential equation for mRNA transcription was formulated for mRNA abundance $S(t)$ in terms of rate parameters for production (α) and degradation (γ) of mRNA:

$$\frac{dS}{dt} = \alpha - \gamma \cdot S(t).$$

This simple model of mRNA production was extended by Zeisel, Köstler, Yarden, and Domany in Zeisel et al. (2011), who added the intermediate step of pre-mRNA processing (with associated parameter β) and included splicing, polyadenylation, and RNA capping (Figure 1). Their new model related abundance of pre-mRNA $U(t)$ with abundance of mature mRNA $S(t)$ as:

$$\frac{dU}{dt} = \alpha - \beta \cdot U(t),$$
$$\frac{dS}{dt} = \beta \cdot U(t) - \gamma \cdot S(t).$$

In this model, mRNA abundance over time, which is represented as dS/dt, is the *velocity* of gene expression. Zeisel et al. (2011) applied their model to microarray data consisting of exon- and intron-specific probes and investigated different gene-specific rates from time series experiments. While microarrays provided genome-wide gene expression measurements, and intron probes could measure abundances of unprocessed transcripts, the technology was noisy and limited in scope (Johnson et al., 2005).

Single-cell RNA sequencing (RNA-seq), which allows for assessing transcript abundances in individual cells by sequencing, is becoming the standard technology with which to probe not only gene expression but also the cellular population structure of organisms (Svensson et al., 2018). Numerous technological breakthroughs continue to reduce the cost per cell. However distinct experiments are still expensive, making it difficult to conduct large-scale single-cell time course experiments. In addition, in many cases, temporal information is limited due to experimental constraints (for example, in studies of human tissues). Nevertheless, single-cell RNA-seq experiments can provide insights into development and can elucidate differentiation relations between cell populations (Cannoodt et al., 2016). "Pseudotime" inference methods (Magwene et al., 2003; Trapnell et al., 2014) infer trajectories of cells by leveraging the fact that in the high-dimensional space of the gene expression, differentiation programs can be modeled as curves connecting progenitor and differentiated cells. While pseudotime methods utilize geometric models of gene expression, they do not connect measurements to underlying molecular kinetics.

In "RNA velocity of single cells" (La Manno et al., 2018), the authors infer parameters for the Zeisel et al. (2011) model from exonic and intronic single-cell RNA-seq data and use them to predict gene expression changes in cells. By making the assumption that processing rates β are the same for all genes (and can therefore all be set to 1), and that mRNA degradation rates are gene-specific constants, they calculate dS/dt from the single-cell RNA-seq data. In addition, the assumptions imply that the ratio of production to degradation (α/γ) is reflected in the data, leading them to a simple expression for velocity of a gene g in a cell t_c:

$$\left[\frac{dS}{dt}\right]_{c,\,g} = U_g(t_c) - \gamma_g \cdot S_g(t_c).$$

This type of analysis can be used infer genes that are being upregulated, are maintaining homeostatic levels, or are being downregulated in individual cells (Figure 1). Additionally, by predicting mRNA levels for all genes at once, "future" cells can be simulated. In the "space" of gene expression, the difference between the predicted cells and the observed cells forms a vector field, providing directions for all cells. These vectors can be used to visualize the kinetic state of all cells in low-dimensional representations of the cell populations.

Long arrows, corresponding to large changes in gene expression, will in the low-dimensional representation imply that cells are undergoing rapid differentiation, while short arrows can be intuitively associated with terminally differentiated cells maintaining homeostasis. While some care needs to be taken when interpreting the lengths of arrows in low-dimensional representations, RNA velocity representations can be biologically informative. In principal component analysis, the vector lengths are meaningful, scaling linearly with RNA velocity estimates in high dimensions. We note that this will not be the case in t-SNE representations (Van der Maaten and Hinton, 2008), though directions remain meaningful and can be interpreted as "RNA arguments" in vector nomenclature.

The study by La Manno et al. illustrates the use of RNA velocity analysis by applying the method to human neurons and progenitors, and elucidating the

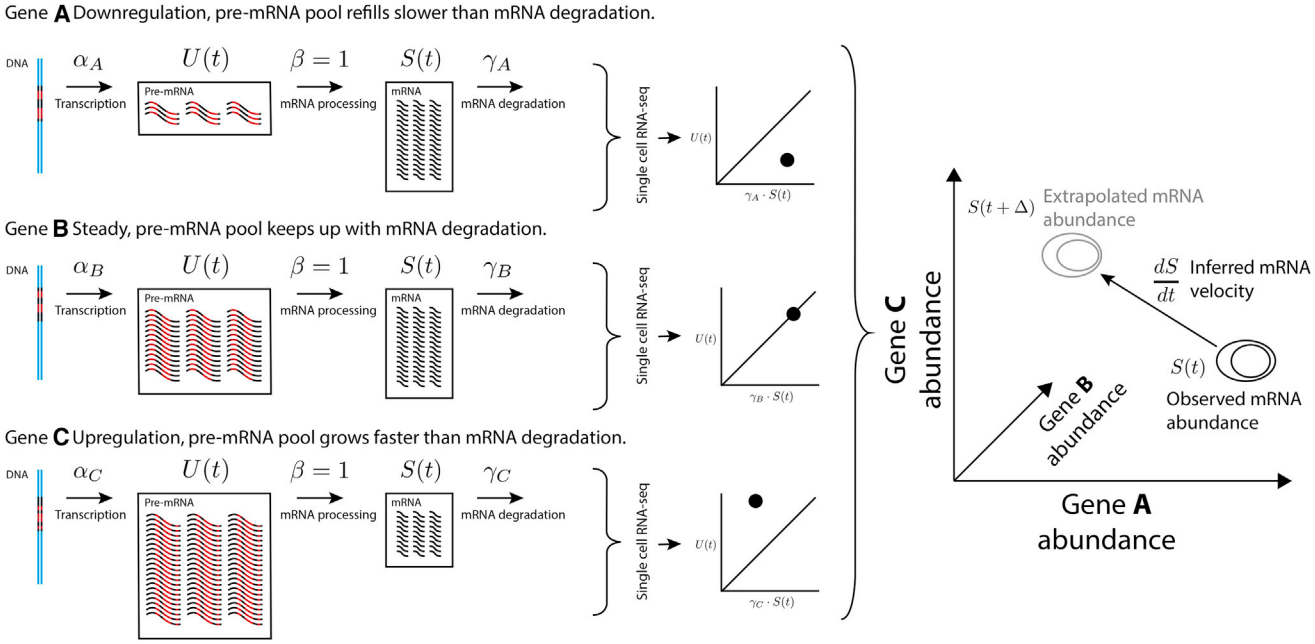

Figure 1. Illustration of the Relation between Pre-mRNA and mRNA in a Single Cell

Three hypothetical genes (A, B, and C) with different kinetic properties. When there are few pre-mRNA molecules compared to mRNA molecules, the mRNA levels will decrease. If the pre-mRNA pool is balanced with mRNA degradation, this indicates the gene will maintain homeostasis. Genes represented by a large pre-mRNA pool are bound to display upregulated mRNA abundances. Both types of mRNA are measured for all genes using single-cell RNA-seq. The ratio between them can be used to predict future mRNA abundance profiles in cells and to extrapolate cell dynamics in gene expression space.

terminal differentiated cell types distinct progenitor cell types are moving toward. Such causal inference has previously only been made on the basis of pairwise similarities between cell types. For example, if neuroblasts on average are closer to immature neurons than mature neurons, one can posit that immature neurons differentiate from neuroblasts into neurons. With RNA velocity, neuroblasts pointing toward immature neurons and immature neurons pointing toward neurons provide a more direct and interpretable analysis. In this way, RNA velocity vectors contribute quantitative evidence grounded in a mechanistic kinetic model of transcription and splicing.

New single-cell genomics protocols are extending single-cell RNA-seq by measuring multiple molecular modalities at the same time in individual cells (Packer and Trapnell, 2018). The RNA velocity approach could be applied to data from protocols such as CITE-seq (Stoeckius et al., 2017) and REAP-seq (Peterson et al., 2017), which measure protein abundance and mRNA abundance in the same cell. Since such technologies also produce information about pre-mRNA and mRNA signals, it may be possible to quantify RNA *acceleration* in single cells using an extension of the Zeisel et al. model. For example, a third equation for change in protein abundance based on a kinetic model of translation (dP/dt) could be included as follows:

$$\frac{dU}{dt} = \alpha - \beta_{\text{processing}} \cdot U(t),$$

$$\frac{dS}{dt} = \beta_{\text{processing}} \cdot U(t) - \gamma_{\text{mRNA}} \cdot S(t),$$

$$\frac{dP}{dt} = \beta_{\text{translation}} \cdot S(t) - \gamma_{\text{protein}} \cdot P(t).$$

The fact that RNA velocities can be computed directly from standard single-cell RNA-seq data means that researchers can use the technique to re-examine previously generated data to answer questions about cell differentiation. Also, future experiments can now be designed with this explicit application in mind without the need to modify optimized protocols. Furthermore, the easy-to-use software implementations provided by La Manno et al. are poised to make RNA velocity analyses standard in single-cell RNA-seq studies.

REFERENCES

Cannoodt, R., Saelens, W., and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. Eur. J. Immunol. *46*, 2496–2506.

Gorini, L., and Maas, W.K. (1957). The potential for the formation of a biosynthetic enzyme in Escherichia coli. Biochim. Biophys. Acta *25*, 208–209.

Johnson, J.M., Edwards, S., Shoemaker, D., and Schadt, E.E. (2005). Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet. *21*, 93–102.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Magwene, P.M., Lizardi, P., and Kim, J. (2003). Reconstructing the temporal ordering of biological samples using microarray data. Bioinformatics *19*, 842–850.

Packer, J., and Trapnell, C. (2018). Single-cell multi-omics: an engine for new quantitative models of gene regulation. Trends Genet. *34*, 653–665.

Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. Nat. Biotechnol. *35*, 936–939.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. Nat. Protoc. *13*, 599–604.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 85.

Zeisel, A., Köstler, W.J., Molotski, N., Tsai, J.M., Krauthgamer, R., Jacob-Hirsch, J., Rechavi, G., Soen, Y., Jung, S., Yarden, Y., and Domany, E. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. Mol. Syst. Biol. *7*, 529.