The differentiation of a cell from an initial to a final state can be viewed as a dynamical process in which the system (i.e., the cell) transits through a sequence of states. This transformation occurs out of equilibrium, typically irreversibly. To characterise how the cell changes during differentiation we can explore modifications in genome architecture, cell morphology or protein content, among others. Here, we describe cellular states using the transcriptome, which is the collection of all RNA molecules present in a cell and can be measured using single-cell RNA sequencing (scRNA-seq). These molecules are transcribed from regions of the DNA known as genes, and encode the sequence of amino acids required to make proteins. By measuring the transcriptome of a cell we can infer which genes it is using and hence what is its main function.

The potential landscape for cell-fate commitment

Seminal work in the field of systems biology has set the ground work to mathematically model the molecular mechanisms underlying cellular lineage commitment. A cell is viewed as a Brownian particle whose state is characterised by a vector of RNA expression, $T = [m_1, m_2, ..., m_N]$, where each dimension is a gene. The probability $P(T,t)$ of a cell to have a transcriptional profile $T$ at time $t$ time satisfies the Fokker-Planck equation:

$$\frac{\partial P(T,t)}{\partial t} = \sum_i \frac{\partial^2}{\partial m_i^2} [D(T)P(T,t)] - \frac{\partial}{\partial m_i} [F_i(T)P(T,t)]$$ (1)

where $F(T) = \{F_1(T), F_2(T), ..., F_N(T)\}$ describes the interactions between genes (known as gene regulatory network, Box), and $D(T)$ represents the stochastic diffusion between states. The solution of Eq. (1) is a high-dimensional
manifold that connects to the Waddington landscape model (Box 1), in which a differentiating cell is represented by a marble that rolls down a landscape made of hills (transition states) and valleys (intermediate states) until reaching the final state (attractor state). We can also describe the cell state with a set of deterministic dynamic equations to describe RNA expression as a function of $F(T)$ and $t$:

$$\frac{\partial m_i}{\partial t} = F_i(T), i = 1, ..., N$$  (2)

A minimalistic regulatory network comprising two interacting molecules can already produce systems displaying bi-/tri-stability depending on the strength of the interactions (see Box). There, a bi-potent progenitor cell can be seen as an attractor state (red solid dot) whose stability is compromised by a change in its self-activation strength. The bi-potent progenitor then becomes a metastable state (red empty dot) that can differentiate towards two distinct terminal fates (cyan dots). This approach has successfully been applied to qualitatively describe lineage commitment in the blood cell types [2]. However, most models have only been experimentally tested using well-defined in vitro differentiation systems where the state of a cell was inferred based on population-averaged ensemble measurements. This reduced the temporal resolution and obscured rare populations and transition states. Hence, a goal in developmental biology is to extract information about the underlying regulatory networks (i.e., $F(T)$) during cell fate commitment in vivo with single-cell resolution and genome-wide.

**Cell type identity and RNA dynamics in single cells**

In multicellular organisms, cells are specialised and perform a limited set of tasks. The identity of a cell (i.e., the cell type) is commonly defined by its function, and the function can be partly determined by its RNA content. Therefore, scRNA-seq experiments provide an excellent platform to identify cell types present in heterogeneous tissues, since they allow measuring the RNA content in thousands of single cells independently (Fig. 1a) [1]. Overall, scRNA-seq provides a snapshot of the cellular heterogeneity in a sample, and can reveal rare or transient subpopulations usually undetectable by bulk sequencing approaches. Because of the huge number of genes in an organism (both the human and the mouse genome have ~20,000 protein-coding genes), datasets are typically large and sparse. It is a priority to develop analytical methods that exploit these high-dimensional data (each dimension being a gene) to compare transcriptome profiles between different cells. At the moment, there are several pipelines to normalise the data, reduce dimensions, cluster similar cells together, extract differential gene expression profiles, visualise results, and Ultimately call cell types (Fig. 1b) [1]. In scRNA-seq, the sample is destroyed when the cells are prepared for sequencing. Therefore, these experiments
are not suitable to measure temporal changes in the transcriptome of the same cell. To circumvent this, experiments are performed on samples where several differentiating cells coexist at different stages (such as embryos), or in biological replicates at different time points, to ensure that all stages of the cellular differentiation trajectory are represented [3].

With the possibility to measure full transcriptomes in single cells undergoing differentiation, new algorithms to investigate the lineage-commitment dynamics have been developed [3]. Usually, it is assumed that cells change their transcriptome continuously as they transit from one state to another. Consequently, cells with similar transcriptomic profiles can be connected together and ordered (Fig. 2a). This gives rise to the concept of pseudo-time as a latent dimension that quantifies the degree of differentiation along a cellular transformation. These approaches are currently challenged by results obtained from lineage tracing studies in which both the transcriptome and the common ancestry of single cells are simultaneously measured (Fig. 2a) [4].

There, it has been observed that transcriptomically identical cells can derive from ancestors with different RNA profiles, suggesting that cell differentiation has a stochastic component in which fluctuations in gene expression play an important role and cells with similar transcriptome can have different lineages.

In order to assign differentiating cells to a pseudo-time value more accurately, experimental protocols and analysis pipelines have been improved to distinguish newly synthesised (u) transcripts from mature (m) ones (Fig. 2b) [5,6]. Then, we can extract the transcription, degradation and splicing rates, \( \alpha \), \( \gamma \) and \( \beta \) respectively, for each transcript, by solving the rate equations:

\[
\frac{du(t)}{dt} = \alpha - \beta u(t) \\
\frac{dm(t)}{dt} = \beta u(t) - \gamma m(t)
\]

(3)

are not suitable to measure temporal changes in the transcriptome of the same cell. To circumvent this, experiments are performed on samples where several differentiating cells coexist at different stages (such as embryos), or in biological replicates at different time points, to ensure that all stages of the cellular differentiation trajectory are represented [3].

With the possibility to measure full transcriptomes in single cells undergoing differentiation, new algorithms to investigate the lineage-commitment dynamics have been developed [3]. Usually, it is assumed that cells change their transcriptome continuously as they transit from one state to another. Consequently, cells with similar transcriptomic profiles can be connected together and ordered (Fig. 2a). This gives rise to the concept of pseudo-time as a latent dimension that quantifies the degree of differentiation along a cellular transformation. These approaches are currently challenged by results obtained from lineage tracing studies in which both the transcriptome and the common ancestry of single cells are simultaneously measured (Fig. 2a) [4].

There, it has been observed that transcriptomically identical cells can derive from ancestors with different RNA profiles, suggesting that cell differentiation has a stochastic component in which fluctuations in gene expression play an important role and cells with similar transcriptome can have different lineages.

In order to assign differentiating cells to a pseudo-time value more accurately, experimental protocols and analysis pipelines have been improved to distinguish newly synthesised (u) transcripts from mature (m) ones (Fig. 2b) [5,6]. Then, we can extract the transcription, degradation and splicing rates, \( \alpha \), \( \gamma \) and \( \beta \) respectively, for each transcript, by solving the rate equations:

\[
\frac{du(t)}{dt} = \alpha - \beta u(t) \\
\frac{dm(t)}{dt} = \beta u(t) - \gamma m(t)
\]

(3)

are not suitable to measure temporal changes in the transcriptome of the same cell. To circumvent this, experiments are performed on samples where several differentiating cells coexist at different stages (such as embryos), or in biological replicates at different time points, to ensure that all stages of the cellular differentiation trajectory are represented [3].

With the possibility to measure full transcriptomes in single cells undergoing differentiation, new algorithms to investigate the lineage-commitment dynamics have been developed [3]. Usually, it is assumed that cells change their transcriptome continuously as they transit from one state to another. Consequently, cells with similar transcriptomic profiles can be connected together and ordered (Fig. 2a). This gives rise to the concept of pseudo-time as a latent dimension that quantifies the degree of differentiation along a cellular transformation. These approaches are currently challenged by results obtained from lineage tracing studies in which both the transcriptome and the common ancestry of single cells are simultaneously measured (Fig. 2a) [4].

There, it has been observed that transcriptomically identical cells can derive from ancestors with different RNA profiles, suggesting that cell differentiation has a stochastic component in which fluctuations in gene expression play an important role and cells with similar transcriptome can have different lineages.

In order to assign differentiating cells to a pseudo-time value more accurately, experimental protocols and analysis pipelines have been improved to distinguish newly synthesised (u) transcripts from mature (m) ones (Fig. 2b) [5,6]. Then, we can extract the transcription, degradation and splicing rates, \( \alpha \), \( \gamma \) and \( \beta \) respectively, for each transcript, by solving the rate equations:

\[
\frac{du(t)}{dt} = \alpha - \beta u(t) \\
\frac{dm(t)}{dt} = \beta u(t) - \gamma m(t)
\]

(3)

are not suitable to measure temporal changes in the transcriptome of the same cell. To circumvent this, experiments are performed on samples where several differentiating cells coexist at different stages (such as embryos), or in biological replicates at different time points, to ensure that all stages of the cellular differentiation trajectory are represented [3].

With the possibility to measure full transcriptomes in single cells undergoing differentiation, new algorithms to investigate the lineage-commitment dynamics have been developed [3]. Usually, it is assumed that cells change their transcriptome continuously as they transit from one state to another. Consequently, cells with similar transcriptomic profiles can be connected together and ordered (Fig. 2a). This gives rise to the concept of pseudo-time as a latent dimension that quantifies the degree of differentiation along a cellular transformation. These approaches are currently challenged by results obtained from lineage tracing studies in which both the transcriptome and the common ancestry of single cells are simultaneously measured (Fig. 2a) [4].

There, it has been observed that transcriptomically identical cells can derive from ancestors with different RNA profiles, suggesting that cell differentiation has a stochastic component in which fluctuations in gene expression play an important role and cells with similar transcriptome can have different lineages.

In order to assign differentiating cells to a pseudo-time value more accurately, experimental protocols and analysis pipelines have been improved to distinguish newly synthesised (u) transcripts from mature (m) ones (Fig. 2b) [5,6]. Then, we can extract the transcription, degradation and splicing rates, \( \alpha \), \( \gamma \) and \( \beta \) respectively, for each transcript, by solving the rate equations:

\[
\frac{du(t)}{dt} = \alpha - \beta u(t) \\
\frac{dm(t)}{dt} = \beta u(t) - \gamma m(t)
\]

(3)

are not suitable to measure temporal changes in the transcriptome of the same cell. To circumvent this, experiments are performed on samples where several differentiating cells coexist at different stages (such as embryos), or in biological replicates at different time points, to ensure that all stages of the cellular differentiation trajectory are represented [3].

With the possibility to measure full transcriptomes in single cells undergoing differentiation, new algorithms to investigate the lineage-commitment dynamics have been developed [3]. Usually, it is assumed that cells change their transcriptome continuously as they transit from one state to another. Consequently, cells with similar transcriptomic profiles can be connected together and ordered (Fig. 2a). This gives rise to the concept of pseudo-time as a latent dimension that quantifies the degree of differentiation along a cellular transformation. These approaches are currently challenged by results obtained from lineage tracing studies in which both the transcriptome and the common ancestry of single cells are simultaneously measured (Fig. 2a) [4].

There, it has been observed that transcriptomically identical cells can derive from ancestors with different RNA profiles, suggesting that cell differentiation has a stochastic component in which fluctuations in gene expression play an important role and cells with similar transcriptome can have different lineages.

In order to assign differentiating cells to a pseudo-time value more accurately, experimental protocols and analysis pipelines have been improved to distinguish newly synthesised (u) transcripts from mature (m) ones (Fig. 2b) [5,6]. Then, we can extract the transcription, degradation and splicing rates, \( \alpha \), \( \gamma \) and \( \beta \) respectively, for each transcript, by solving the rate equations:

\[
\frac{du(t)}{dt} = \alpha - \beta u(t) \\
\frac{dm(t)}{dt} = \beta u(t) - \gamma m(t)
\]

(3)
It has been proved that the experimental estimation of the time derivative of the mature RNA —termed RNA velocity— can predict the future state of individual cells on a timescale of hours [5]. RNA velocity can be embedded in the two dimensional representation of scRNA-seq data, in the form of a phase portrait from which one can visualise directions of differentiation at the cellular level (Fig. 2c).

A new challenge of developmental biology is to experimentally connect Eq. (2) and Eq. (3), and link RNA velocity to F(T) to unravel the regulatory networks that drive cell-fate commitment. This will allow experimentally measuring the Waddington landscape, with the goal to ultimately predict cellular responses to stimuli.

Future perspectives: lineage tracing, spatial transcriptomics and perturbation assays

The characterisation of gene regulatory networks and cellular transition states occurring during cell differentiation and their role in cell-fate commitment is a long-standing question in developmental biology. We anticipate that the combination of RNA velocity with theoretical models of molecular phase portraits will help us unravel the fundamental gene regulatory networks that give rise to new cell fates. With the help of lineage tracing and perturbation studies (such as chemical screens or genetic knock-outs), we will be able to validate our findings and test predictions. In combination with spatial transcriptomics [7], we expect to also unravel underlying principles of intracellular communication.

The ongoing revolution in single-cell sequencing technologies is making it possible to extract other epigenetic properties in single cells, and to integrate them into multi-modal measurements from the same cell. With this, we expect to enhance the characterization of the Waddington landscape of development, with all its epigenetic components.

About the Author

Anna Alemany received her PhD in the group of Prof. Ritort and afterwards joined Prof. van Oudenaarden laboratory for her postdoc. She will start her own laboratory in the Leiden Medical University Center, in the Netherlands. Her goal is to quantitatively characterise the Waddington landscape of development using single-cell RNA sequencing experiments.

References