dyngen: benchmarking with in silico single cells

Robrecht Cannoodt Wouter Saelens Louise Deconinck Yvan Saeys

09 December 2019

Purpose: A common problem of pioneering computational tools is that during their development, there are rarely sufficient datasets available for adequately quantitatively assessing its performance.
Results: We developed dyngen, a multi-modality simulator of single cells. In dyngen, the biomolecular state of an *in silico* cell changes over time according to a set of reactions defined by the cell's gene regulatory network. By simulating single cells in terms of its biomolecular state and reactions, the simulator is easily extendible for adding new modalities or experimental procedures. We demonstrate dyngen's flexibility by simulating snapshot, time-series and perturbation experiments.
Conclusion: dyngen lays the foundations for benchmarking a wide variety of computational single-cell tools and can be used to help kick-start the development of future types of analyses.

Author contributions:

- W.S. and R.C. designed the study.
- R.C., W.S., and L.D. performed the experiments and analysed the data.
- R.C. and W.S. implemented the dyngen software package.
- R.C. and W.S. wrote the original manuscript.
- L.D. wrote the section on trajectory alignment.
- R.C., W.S., L.D., and Y.S. reviewed and edited the manuscript.
- Y.S. supervised the project.

Introduction

Continuous technological advancements to single-cells omics are having profound effects on how researchers can validate biological hypotheses. Early experimental technologies typically only allowed profiling a single modality (e.g. DNA sequence, RNA or protein expression). However, recent developments permit profiling multiple modalities simultaneously, and every modality added allows for new types of analyses that can be performed.

This presents method developers with a problem. The majority of the 250+ peer-reviewed computational tools for analysing single cell omics data were published without a quantitative assessment of the accuracy of the tool. This is partially due to low availability of suitable benchmarking datasets; even if there are sufficient suitable input datasets available, these are often not accompanied by the necessary metadata to serve as ground-truth for a benchmark.

Here, synthetic data plays a crucial role in asserting minimum performance requirements for novel tools in anticipation of adequate real data. Generators of scRNA-seq data (e.g. splatter[1], powsimR[2], PROSSTT[3] and SymSim[4]) have already been widely used to explore the strengths and weaknesses of computational tools, both by method developers[5, 6, 7, 8] and independent benchmarkers[9, 10, 11]. However, a limitation of scRNA-seq profiles generators is that they would require significant methodological alterations to add additional modalities or experimental conditions.

An ideal experiment would be able to observe all aspects of a cell, including a full history of its molecular states, spatial positions and environmental interactions[12]. While this falls outside the reach of current experimental technologies, generating synthetic data in anticipation of new experimental technologies would allow already developing the next wave of computational tools.

We developed a multi-modality simulator of single cells called dyngen (Figure 1A). dyngen uses Gillespie's stochastic simulation algorithm[13] to simulate gene regulatory interactions at a single-molecule level. Its methodology allows tracking of many layers of information throughout the simulation, including the abundance of any molecule in the cell, progression of the cell along a dynamic process, and the activation strength of individual regulatory interactions. dyngen can simulate a large variety of dynamic processes (e.g. cyclic, branching, disconnected) as well as a broad

range of experimental conditions (e.g. batch effects and time-series, perturbation and knockdown experiments). The fine-grained controls over simulation parameters allow dyngen to be applicable to a broad range of use-cases. We demonstrate this by performing first quantitative evaluations of three types of novel computational approaches: RNA velocity, casewise network inference and trajectory alignment methods.

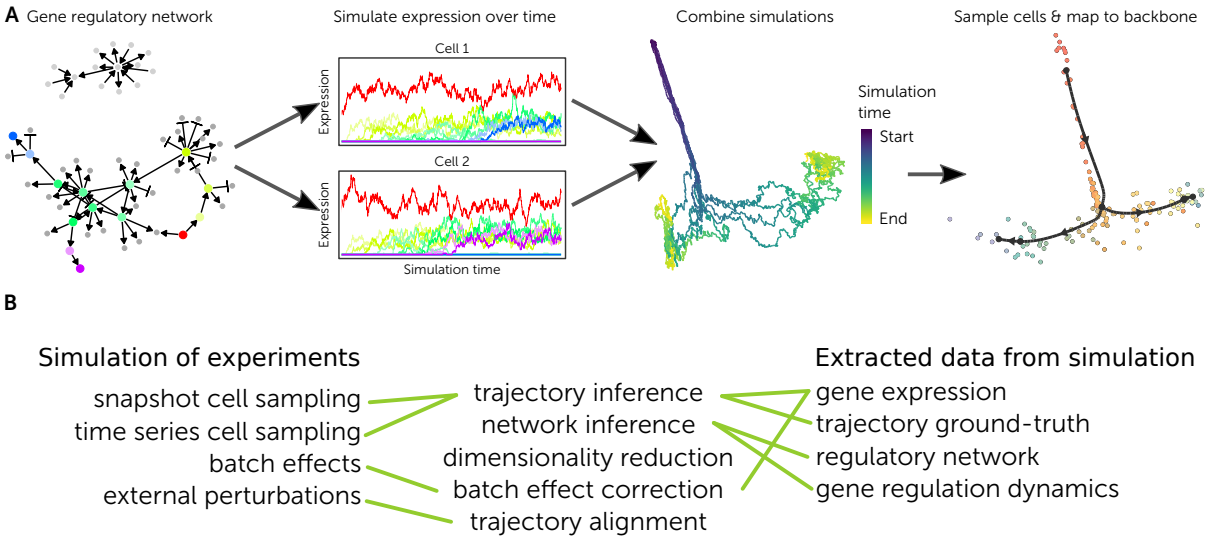

Figure 1: **Showcase of dyngen functionality. A:** The typical process of generating a dataset with dyngen. **B:** Evaluating different types of computational tools requires simulating different types of experiments and extracting different layers of information from the simulation.

Results

A cell consists of a set of molecules, the abundance of which are affected by a set of reactions: transcription, splicing, translation, and degradation (Figure 2A). A gene regulatory network (GRN) defines the reactions that are allowed to occur (Figure 2B), which is constructed in such a way that cells slowly develop over time (Figure 2C,D). With every time step dt in the simulation, the probability of a reaction occurring is computed (not shown). From the probabilities are sampled which reactions occur during this time step dt (Figure 2E).

dyngen returns many modalities throughout the whole simulation: molecular abundance, cellular state, number of reaction firings, reaction likelihoods, and regulation activations (Figure 2C–F). These modalities can serve both as input data and ground truth for benchmarking many types of computational approaches. For example, a network inference method could use mRNA abundance and cellular states as inputs and its output could be benchmarked against the gold standard GRN.

Depending on how the GRN is designed, different cellular developmental processes can be simulated. dyngen includes generators of GRNs which result in many different developmental topologies (Figure 3), including branching, converging, cyclic and even disconnected. Custom-defined GRNs offer more fine-grained control over the simulation.

Together, these qualities allow it to be applicable in benchmarking a broad range of use-cases. To demonstrate, we apply dyngen on several novel computational approaches which have recently had a major impact on how single-cell analyses are performed but for which quantitative assessment of the performance was hitherto lacking.

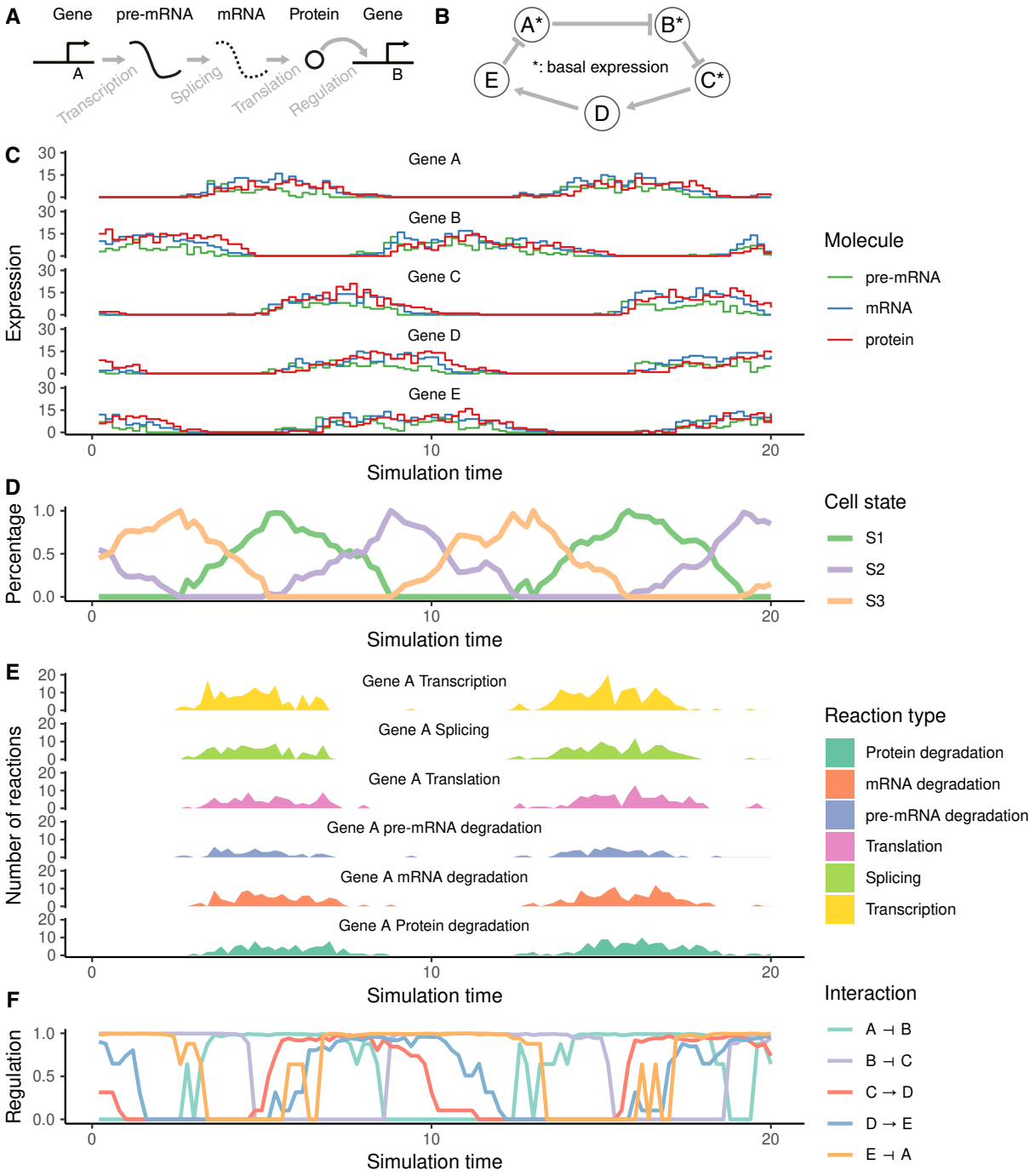

Figure 2: **dyngen models reactions at a single-molecule level and keeps track of multiple levels of information throughput a simulation. A:** Changes in abundance levels are driven strictly these gene regulatory reactions. **B:** The input GRN is defined such that it models a dynamic process of interest. **C:** The reactions define how abundance levels of molecules change at any particular timepoint. **D:** Firing many reactions can significantly alter the cellular state over time. **E:** dyngen keeps track of the reactions that were fired during small intervals of time. **F:** Similarly, dyngen can also keep track of the regulatory activity of every interaction.

Example use cases

Trajectory alignment

Casewise network inference

Casewise network inference (CNI) methods[1] predict not only which transcription factors regulate which target genes (Figure 4A, top left), but also how active each interaction is in every case (Figure 4A). In CNI, a 'case' might be a cell,

[1]Other terms are commonly used when dealing with data from a particular source. For example, single-cell NI when applied to single-cell transcriptomics data; sample-specific NI when applied to bulk transcriptomics; patient-derived NI when applied to bulk profiles of patients.

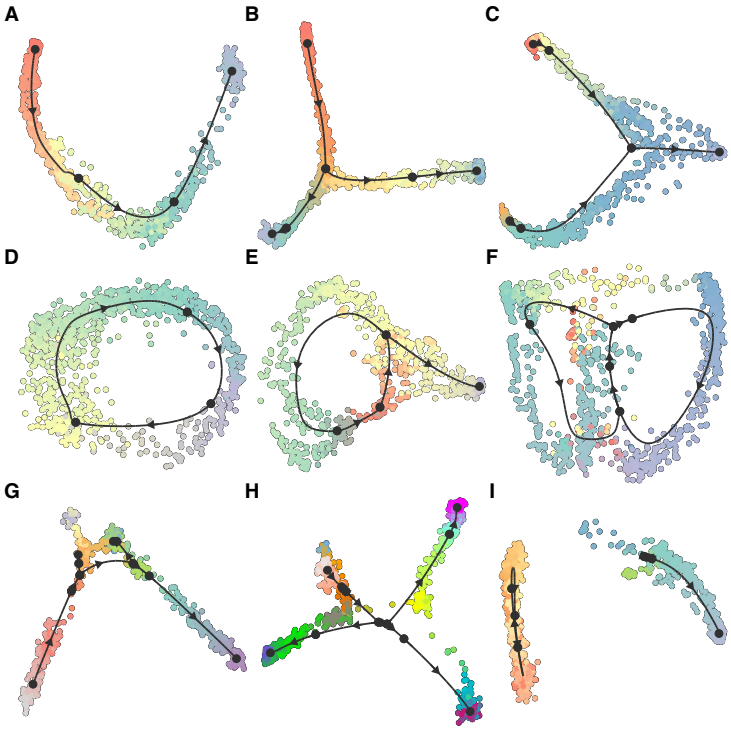

Figure 3: **Multiple executions of dyngen with different predefined backbones.** From each simulation of about 200 genes, 1000 cells were sampled. **A:** Linear. **B:** Bifurcating. **C:** Converging. **D:** Cyclic. **E:** Bifurcating loop. **F:** Bifurcating converging. **G:** Consecutive branching. **H:** Binary tree. **I:** Disconnected.

but it might also refer to a bulk sample.

While a few pioneering CNI approaches have already been developed[14, 15, 16], a quantitative assessment of the performance is hitherto lacking. This is not surprising, as neither real nor in silico datasets of cell-specific or even cell-type-specific interactions exists that is large enough so that it can be used as a ground-truth for evaluating CNI methods.

Since dyngen computes tracks the probability of transcription, temporarily 'knocking down' the expression of a regulator and observing the change in transcription probability. This is a much more accurate ground-truth for regulatory interaction between a regulator and target in comparison to observing the change in transcript abundance levels when knocking down a regulator, as the regulation of the target could be indirect.

We used this ground-truth to compare the performance of three CNI methods (Figure 4B). We calculated the AUROC and AUPR score – which are common metrics for NI benchmarking – for each cell individually. Computing the mean AUROC and AUPR per dataset showed that pySCENIC significantly outperforms LIONESS and SSN.

This comparison could be extended to include analyses on the scalability of execution time w.r.t dataset size, the stability of the results in function of noise, and the usability toward end users.

RNA velocity

In eukaryotes, a gene is first transcribed to a pre-mRNA, and subsequently spliced into mature mRNA. Because reads coming from both unspliced and spliced transcripts are observed in expression data, the relative ratio between the two can tell us something about which genes are increasing, decreasing or remaining the same[17, 18]. To determine this, some parameters have to be estimated to determine which fraction of unspliced and spliced mRNAs corresponds to an increase or decrease. The estimation of these parameters makes some assumptions, and can be handled in different ways in the two main algorithms that are now available for RNA velocity estimation: velocyto[18] and scvelo[19]. It can be difficult to obtain ground truth data to benchmark these algorithms, given that it would require continuous data of transcriptional dynamics in individual cells. On the other hand, the ground truth velocity is rapidly extracted from the dyngen model, by looking whether each transcript is currently increasing or decreasing in expression.

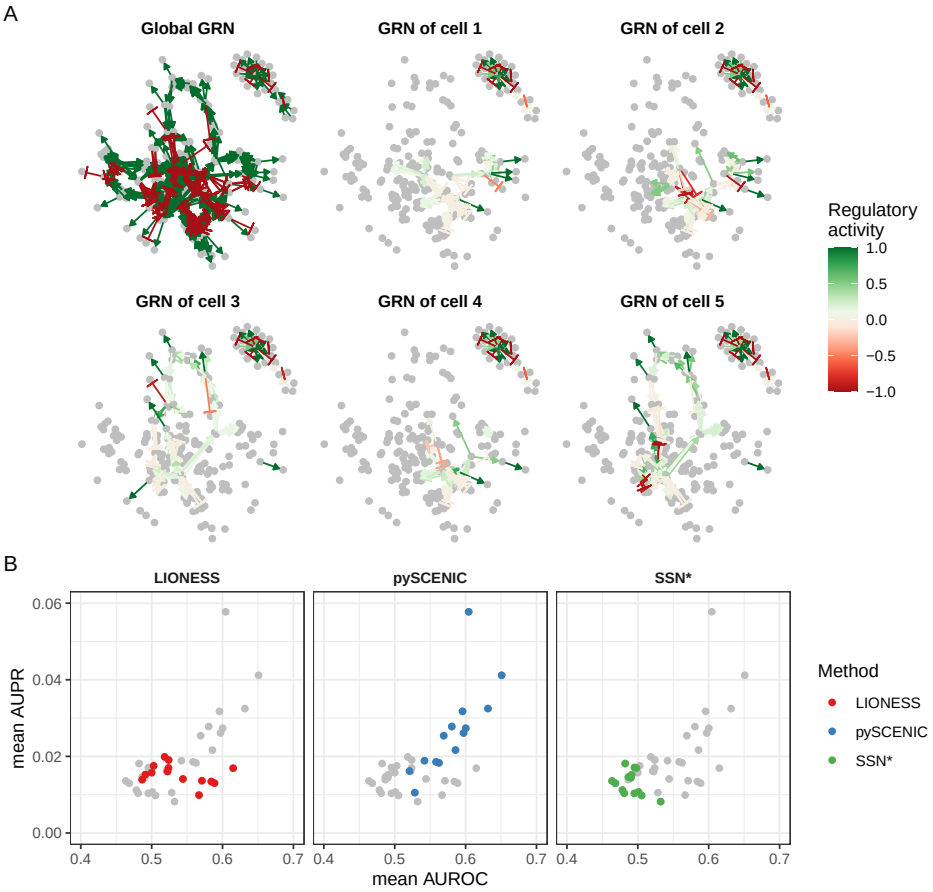

Figure 4: **dyngen allows benchmarking Casewise Network Inference (CNI) methods. A:** A cell is simulated using the global gene regulatory network (GRN, top left). However, at any particular state in the simulation, only a fraction of the gene regulatory interactions are active. **B:** CNI methods were executed to predict the regulatory interactions that are active in each cell specifically. Using the ground-truth casewise GRN, the performance of each method was quantified on 14 dyngen datasets.

We tested scvelo and velocyto on 8 datasets containing linear, bifurcating, disconnected and cyclic trajectories, and varied the main parameter settings in which they estimate the velocity. We found that the original velocyto implementation, which assumes that the velocity remains constant in some cells, performed the best across all datasets. The dynamical estimation of velocyto, as implemented in scvelo, performed the worst of all parameter settings. This was mainly due to scvelo overestimating the dynamics of a gene, especially towards upregulation, while velocyto correctly estimated not only when a gene changes, but also when it remained in a steady state.

Discussion

As is, dyngen's single cell simulations can be used to evaluate common single-cell omics computational methods such as clustering, batch correction, trajectory inference and network inference. However, the combined effect of these advantages results in a framework that is flexible enough to adapt to a broad range of applications. This may include methods that integrate clustering, network inference and trajectory inference. In this respect, dyngen may promote the development of new tools in the single-cell field similarly as other simulators have done in the past[20, 21].

dyngen ultimately allows anticipating technological developments in single-cell multi-omics. In this way, it is possible to design and evaluate the performance and robustness of new types of computational analyses before experimental data becomes available. In addition, it could also be used to compare which experimental protocol is the most cost-effective in producing the qualitative and robust results in downstream analysis.

Currently, dyngen focuses on simulating cells as standalone entities that are well mixed. Splitting up the simula-

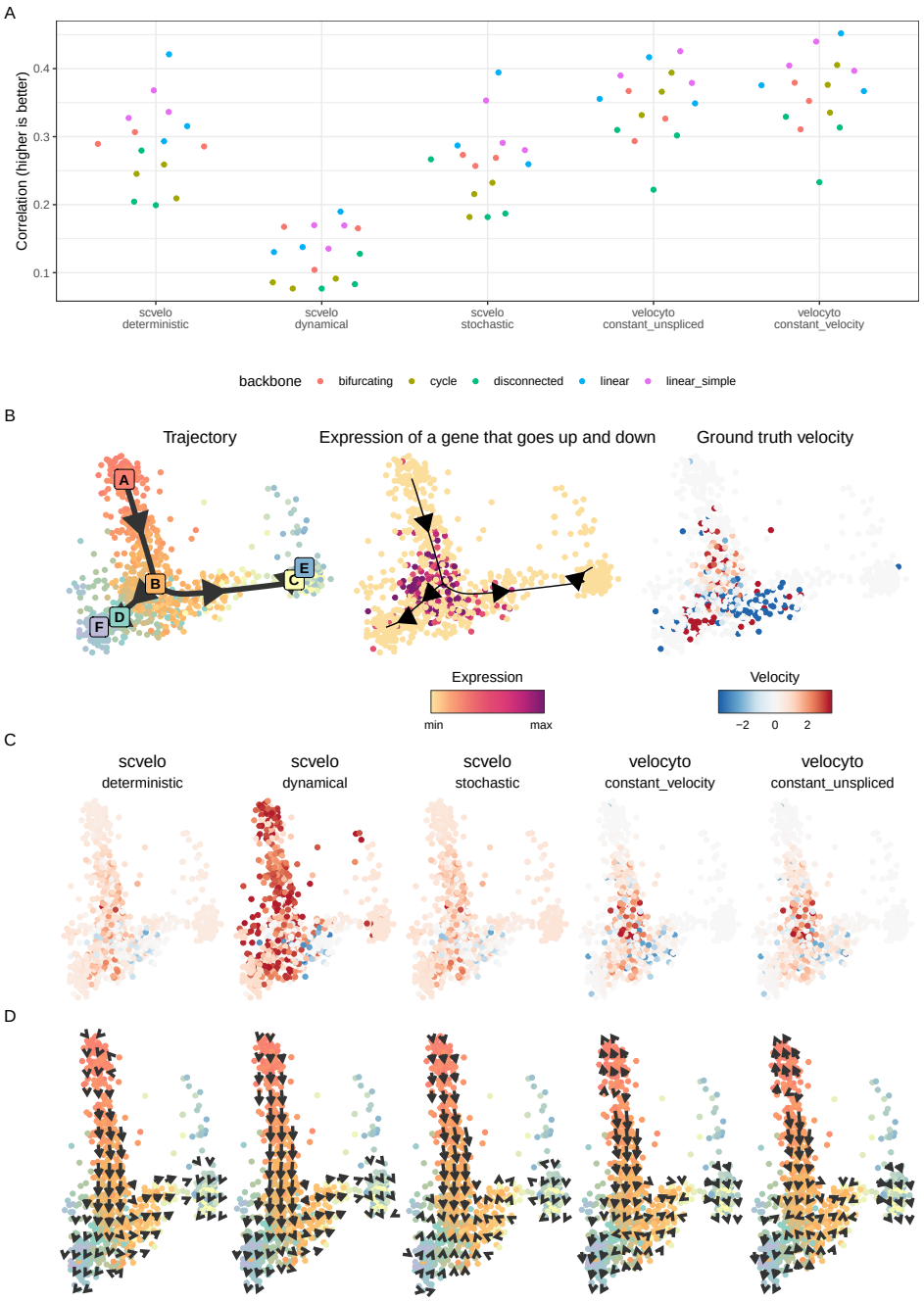

Figure 5: **dyngen allows benchmarking of RNA velocity methods. A:** We tested five different methods and parameter settings for the estimation of RNA velocity on datasets with varying backbones (colors). Overall, the velocyto method with the constant velocity assumption performed the best overall. **B:** An example bifurcating dataset, with as illustration the expression and ground truth velocity of a gene that goes up and down in the trajectory. **C:** The RNA velocity estimates of the different methods. **D:** The embedded RNA velocity of the different methods.

tion space into separate subvolumes could pave the way to better study key cellular processes such as cell division, intercellular communication and migration[22].

Methods

The workflow to generate *in silico* single cell data consists of six main steps (Figure 6).

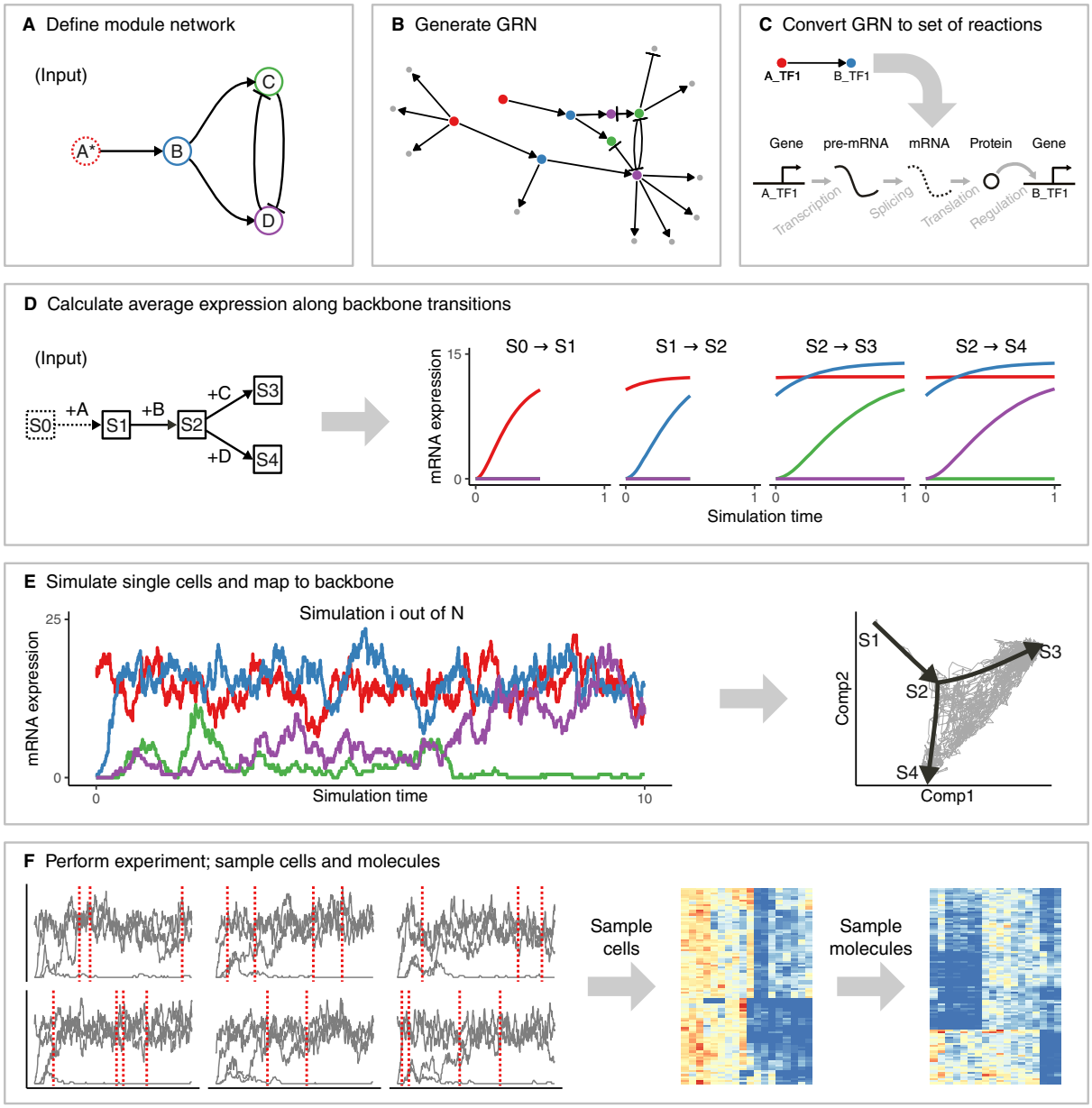

Figure 6: **The workflow of dyngen is comprised of six main steps. A:** The user needs to specify the desired module network or use a predefined module network. **B:** Each gene in a module is is regulated by one or more transcription factors from the upstream module. Additional target genes are generated. **C:** Each gene regulatory interaction in the GRN is converted to a set of biochemical reactions. **D:** Along with the module network, the user also needs to specify the backbone structure of expected cell states. The average expression of each edge in the backbone is simulated by activating a restricted set of genes for each edge. **E:** Multiple Gillespie SSA simulations are run using the reactions defined in step C. The counts of each of the molecules at each time step are extracted. Each time step is mapped to a point in the backbone. **F:** The molecule levels of multiple simulations are shown over time (left). From each simulation, multiple cells are sampled (from left to middle). Technical noise from profiling is simulated by sampling molecules from the set of molecules inside each cell (from middle to right).

Defining the backbone: modules and states

One of the main processes involved in cellular dynamic processes is gene regulation, where regulatory cascades and feedback loops lead to progressive changes in expression and decision making. The exact way a cell chooses a certain path during its differentiation is still an active research field, although certain models have already emerged and been tested *in vivo*. One driver of bifurcation seems to be mutual antagonism, where two genes strongly repress each other[23, 24], forcing one of the two to become inactive[25]. Such mutual antagonism can be modelled and

simulated[26, 27]. Although the two-gene model is simple and elegant, the reality is frequently more complex, with multiple genes (grouped into modules) repressing each other[28].

To start a dyngen simulation, the user needs to define a module network and a backbone. The module network defines how sets of co-regulated genes, called modules, regulate each other. The module network is what mainly determines which dynamic processes occur within the simulated cells. The backbone is a separate set of simulations in which the ground-truth topology of the dynamic processes are defined, as it is difficult to determine the topology of the dynamic processes from the module network itself.

A module network consists of modules connected together by regulatory interactions. A module may have basal expression, which means genes in this module will be transcribed without the presence of transcription factor molecules. A module marked as "active during the burn phase" means that this module will be allowed to generate expression of its genes during an initial warm-up phase (See section). At the end of the dyngen process, cells will not be sampled from the burn phase simulations. Interactions between modules have a strength (which is a positive integer) and an effect (+1 for upregulating, -1 for downregulating).

Several examples of module networks are given (Figure 7). A simple chain of modules (where one module upregulates the next) results in a *linear* process. By having the last module repress the first module, the process becomes *cyclic*. Two modules repressing each other is the basis of a *bifurcating* process, though several chains of modules have to be attached in order to achieve progression before and after the bifurcation process. Finally, a *converging* process has a bifurcation occurring during the burn phase, after which any differences in module regulation is removed.

Note that these examples represent the bare minimum in terms of number of modules used. Using longer chains of modules is typically desired. In addition, the fate decisions made in this example of a bifurcation is reversible, meaning cells can be reprogrammed to go down a different differentiation path. If this effect is undesirable, more safeguards need to be put in place to prevent reprogramming from occurring (Section).

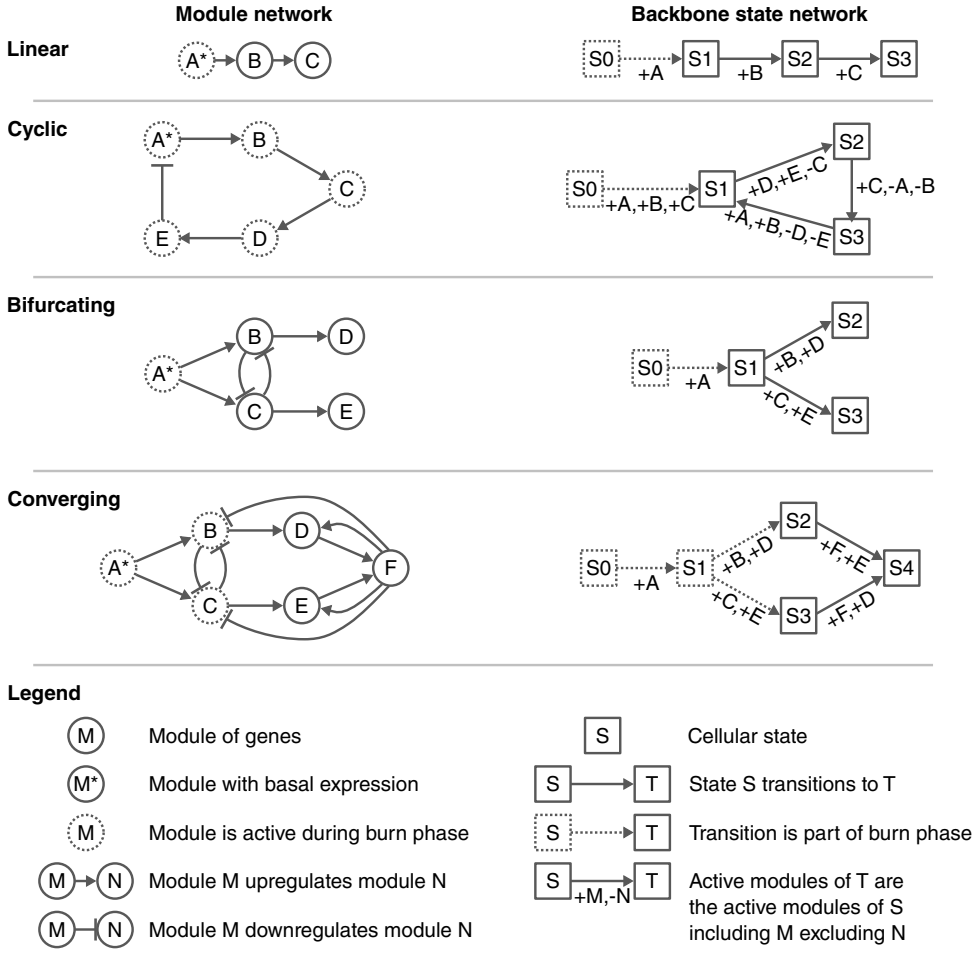

Figure 7: **Example module networks.**

In addition to the module network, the user also needs to define a network of cellular states called the "backbone".

Before simulating any cells, each transition in the backbone is simulated separately to obtain the average changes in expression along that transition (Figure 6D). As part of the backbone, the user needs to specify which modules are allowed to alter its expression from one state to another. For example, in order to transition from state S0 to S1 in the cyclic example, gene modules A, B and C are turned on and a simulation is allowed to run. To transition from S1 to S2, gene modules D and E are turned on, and expression of gene module C is kept constant. To transition from S2 to S3, C is turned on again and now A and B are fixed. Finally, to transition from S3 to S1 again, A and B are turned on again and D and E are fixed again. Demonstrations of the backbone will be explained in more detail in section .

Backbone lego

The backbone can make use of one or more "backbone lego" (BBL) pieces (Figure 8). A BBL consists of one or more modules which regulate each other such that the output modules present a specific behaviour, depending on the input module (Figure 8A). Parameters allow determining the number of modules involved in the process and the number of outputs. Multiple BBLs can be chained together in order to intuitively create module networks and corresponding state networks (Figure 8B). Note that not all dynamic processes can be represented by a combination of BBLs, but they can serve as common building blocks to aid the construction of the backbone.

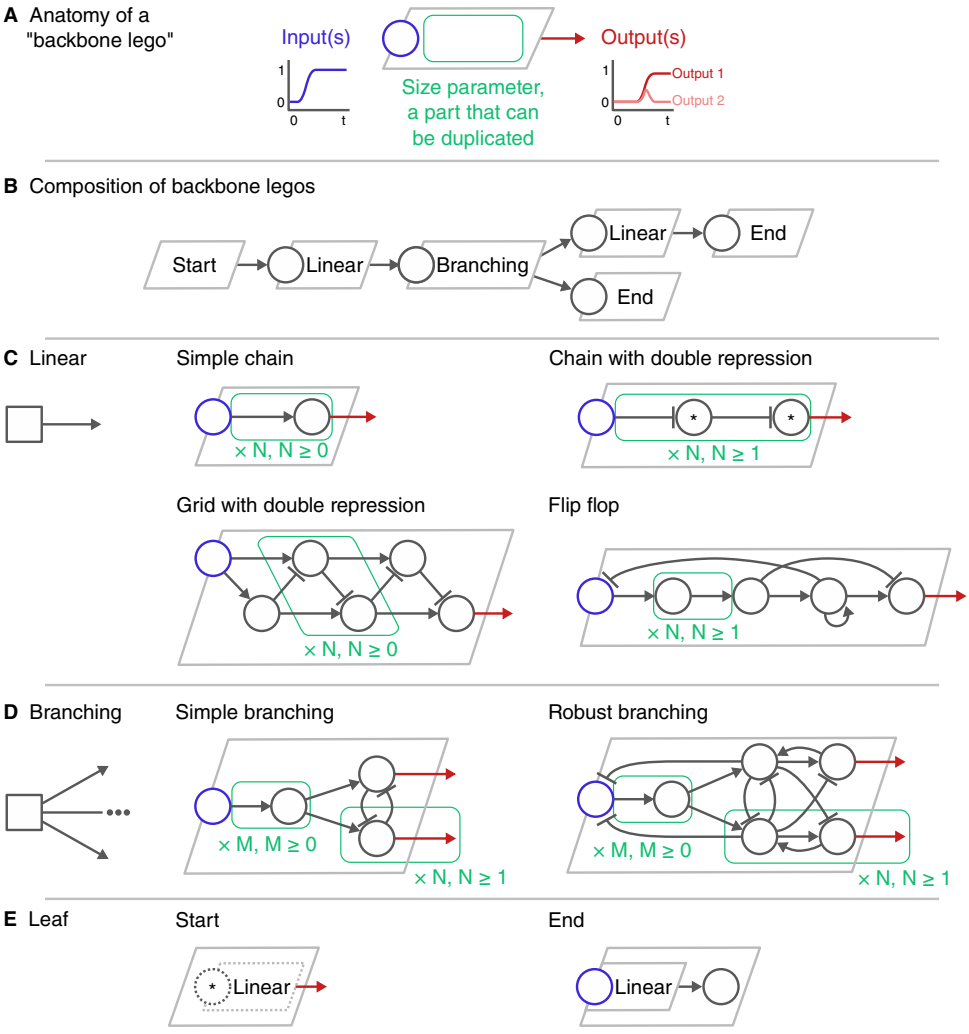

Figure 8: **Backbone lego.**

When the input node of a **linear BBL** (Figure 8C) is upregulated, the module the BBL is connected to will be upregulated. A *simple chain* is a set of modules where a module upregulates the next. A *chain with double repression* has an uneven number of modules forming a chain where each module downregulates the next but all modules (except the input) have basal expression. A *grid with double repression* is similar; except that modules do not have basal expression but instead get upregulated by an upstream module in the chain. Finally, a *flip flop*} consists of a simple chain where first the modules (except the last) are upregulated. Once the second to last module is upregulated, that

module upregulates itself and the first module is strongly repressed, causing all other modules to lose expression and finally the last module to be upregulated. The *flip flop* retains this output state, even when the input changes.

When the input node of a **branching BBL** (Figure 8D) is upregulated, a subset of its output modules will eventually be upregulated. A *simple branching* uses reciprocal inhibition to drive the upregulation of one of the output modules. Due to its simplicity, however, multiple output modules might be upregulated simultaneously and over long periods of simulation time it might be possible that the choice of upregulated module changes. A *robust branching* improves upon the simple branching by preventing upregulation of output modules until an internal branching decision has been made, and by repressing the decision mechanism to avoid other output modules being upregulated other than the one that has been chosen.

A **leaf BBL** (Figure 8E) is a linear BBL that has either no inputs or no outputs. A *start* BBL is a linear BBL where the first module has basal expression, and all modules in this module will be active during the burn-in phase of the simulation (Section). An *end* BBL is also a linear BBL with its output regulating one final module.

Generating the gene regulatory network

The GRN is generated based on the given backbone in four main steps (Figure 9).

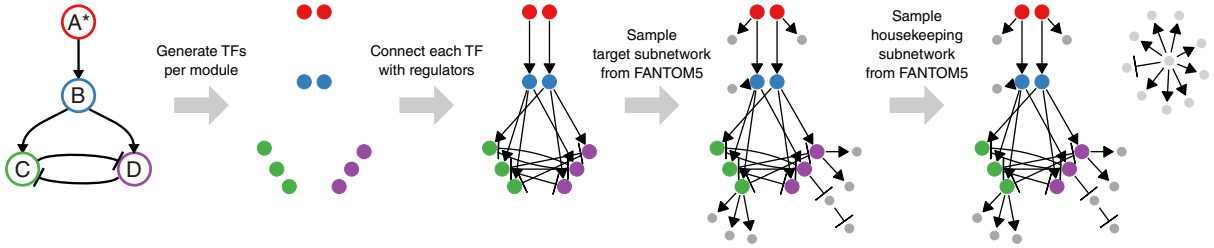

Figure 9: **Generating the feature network from a backbone consists of four main steps.**

Step 1, sampling the transcription factors (TF). The TFs are the main drivers of the molecular changes in the simulation. The user provides a backbone and the number of TFs to generate. Each TF is assigned to a module such that each module has at least x parameters (default $x = 1$). A TF inherits the 'burn' and 'basal expression' from the module it belongs to.

Step 2, generating the TF interactions. Let each TF be regulated according to the interactions in the backbone. These interactions inherit the effect, strength, and cooperativity parameters from the interactions in the backbone. A TF can only be regulated by other TFs or itself.

Step 3, sampling the target subnetwork. A user-defined number of target genes are added to the GRN. Target genes are regulated by a TF or another target gene, but is always downstream of at least one TF. To sample the interactions between target genes, one of the many FANTOM5 [29] GRNs is sampled. The currently existing TFs are mapped to regulators in the FANTOM5 GRN. The targets are drawn from the FANTOM5 GRN, weighted by their page rank value. For each target, at most x regulators are sampled from the induced FANTOM5 GRN (default $x = 5$). The interactions connecting a target gene and its regulators are added the GRN.

Step 4, sampling the housekeeping subnetwork. Housekeeping genes are completely separate from any TFs or target genes. A user-defined set of housekeeping genes are also sampled from the FANTOM5 GRN. The interactions of the FANTOM5 GRN are first subsampled such that the maximum in-degree of each gene is x (default $x = 5$). A random gene is sampled and a breadth-first-search is performed to sample the desired number of housekeeping genes.

Convert gene regulatory network to a set of reactions

Simulating a cell's GRN makes use of a stochastic framework which tracks the abundance levels of molecules over time in a discrete quantity. For every gene G, the abundance levels of three molecules are tracked, namely of corresponding pre-mRNAs, mature mRNAs and proteins, which are represented by the terms w_G, x_G and y_G respectively. The GRN defines how a reaction affects the abundance levels of molecules and how likely it will occur. Gibson and Bruck[30] provide a good introduction to modelling gene regulation with stochastic frameworks, on which many of the concepts below are based.

For every gene in the GRN a set of reactions are defined, namely transcription, splicing, translation, and degradation. Each reaction consists of a propensity function – a formula $f(.)$ to calculate the probability $f(.) \times \mathrm{d}t$ of it occurring during a time interval $\mathrm{d}t$ – and the effect – how it will affect the current state if triggered.

The effects of each reaction mimic the respective biological processes (Table 1, middle). Transcription of gene G results in the creation of a single pre-mRNA molecule w_G. Splicing turns one pre-mRNA w_G into a mature mRNA w_G. Translation uses a mature mRNA x_G to produce a protein y_G. Pre-mRNA, mRNA and protein degradation results in the removal of a w_G, x_G, and y_G molecule, respectively.

The propensity of all reactions except transcription are all linear functions (Table 1, right) of the abundance level of some molecule multiplied by a constant drawn from a normal distribution (Table 2). The propensity of transcription of a gene G depends on the abundance levels of its TFs.

Table 1: **Reactions affecting the abundance levels of pre-mRNA w_G, mature mRNA x_G and proteins y_G of gene G.** Define the set of regulators of G as R_G, the set of upregulating regulators of G as R_G^+, and the set of downregulating regulators of G as R_G^-. Parameters used in the propensity formulae are defined in Table 2.

Reaction	Effect	Propensity
Transcription	$\emptyset \to \mathsf{w}_G$	$\mathsf{wpr}_G \times \frac{\mathsf{ba}_G - \mathsf{co}_G^{\lvert\mathsf{R}_G^+\rvert} + \prod_{H \in \mathsf{R}_G^+} (\mathsf{co}_G + \chi_{G,H})}{\prod_{H \in \mathsf{R}_G} (1 + \chi_{G,H})}$
Pre-mRNA degradation	$\mathsf{w}_G \to \emptyset$	$\mathsf{wdr}_G \times \mathsf{w}_G$
Splicing	$\mathsf{w}_G \to \mathsf{x}_G$	$\mathsf{wsr}_G \times \mathsf{w}_G$
Mature mRNA degradation	$\mathsf{x}_G \to \emptyset$	$\mathsf{xdr}_G \times \mathsf{x}_G$
Translation	$\mathsf{x}_G \to \mathsf{x}_G + \mathsf{y}_G$	$\mathsf{ypr}_G \times \mathsf{x}_G$
Protein degradation	$\mathsf{y}_G \to \emptyset$	$\mathsf{ydr}_G \times \mathsf{y}_G$

Table 2: **Default parameters defined for the calculation of reaction propensity functions.**

Parameter	Symbol	Definition
Transcription rate	wpr_G	$\in N(50, 10),\ \geq 10$
Splicing rate	wsr_G	$\in N(5, 1),\ \geq 1$
Translation rate	ypr_G	$\in N(5, 1),\ \geq 1$
Pre-mRNA half-life	whl_G	$\in N(0.15, 0.03),\ \geq 0.05$
Mature mRNA half-life	xhl_G	$\in N(0.15, 0.03),\ \geq 0.05$
Protein half-life rate	yhl_G	$\in N(0.25, 0.05),\ \geq 0.1$
Interaction strength	$\mathsf{str}_{G,H}$	$\in 10^{U(0,2)}$ *
Hill coefficient	$\mathsf{hill}_{G,H}$	$\in U(0.5, 2)$ *
Cooperativity factor	co_G	$\in [0, 1]$ *
Pre-mRNA degradation rate	wdr_G	$= \ln(2) \,/\, \mathsf{whl}_G$
Mature mRNA degradation rate	xdr_G	$= \ln(2) \,/\, \mathsf{xhl}_G$
Protein degradation rate	ydr_G	$= \ln(2) \,/\, \mathsf{yhl}_G$
Dissociation constant	dis_H	$= 0.5 \times \frac{\mathsf{wpr}_H \times \mathsf{wsr}_H \times \mathsf{ypr}_H}{(\mathsf{wdr}_H + \mathsf{wsr}_H) \times \mathsf{xdr}_H \times \mathsf{ydr}_H}$
Binding	$\chi_{G,H}$	$= \left(\mathsf{str}_{G,H} \times \mathsf{y}_H \,/\, \mathsf{dis}_H\right)^{\mathsf{hill}_{G,H}}$
Basal expression	ba_G	$= \begin{cases} 1 & \text{if } \mathsf{R}_G^+ = \emptyset \\ 0.0001 & \text{if } \mathsf{R}_G^- = \emptyset \text{ and } \mathsf{R}_G^+ \neq \emptyset \text{ *} \\ 0.5 & \text{otherwise} \end{cases}$

*: unless G is a TF, then the value is determined by the backbone.

The propensity of the transcription of a gene G is inspired by thermodynamic models of gene regulation [need-citation], in which the promoter of G can be bound or unbound by a set of N transcription factors H_i. Let $f(\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_N)$ denote the propensity function of G, in function of the abundance levels of the transcription factors. The following subsections explain and define the propensity function when $N = 1$, $N = 2$, and finally for an arbitrary N.

Propensity of transcription when $N = 1$

In the simplest case when $N = 1$, the promoter can be in one of two states. In state S_0, the promoter is not bound by any transcription factors, and in state S_1 the promoter is bound by H_1. Each state S_j is linked with a relative activation α_j, a number between 0 and 1 representing the activity of the promoter at this particular state. The propensity function is thus equal to the expected value of the activity of the promoter multiplied by the pre-mRNA production rate of G.

$$f(y_1, y_2, \dots, y_N) = \text{wpr} \cdot \sum_{j=0}^{2^N-1} \alpha_j \cdot P(S_j) \tag{1}$$

(2)

For $N = 1$, $P(S_1)$ is equal to the Hill equation, where k_i represents the concentration of H_i at half-occupation and n_i represents the Hill coefficient. Typically, n_i is between [1,10]

$$P(S_1) = \frac{y_1^{n_1}}{k_1^{n_1} + y_1^{n_1}} \tag{3}$$

$$= \frac{(y_1/k_1)^{n_1}}{1 + (y_1/k_1)^{n_1}} \tag{4}$$

The Hill equation can be simplified by letting $\nu_i = \left(\frac{y_i}{k_i}\right)^{n_i}$.

$$P(S_1) = \frac{\nu_1}{1 + \nu_1} \tag{5}$$

Since $P(S_0) = 1 - P(S_1)$, the activation function is formulated and simplified as follows.

$$f(y_1) = \text{wpr} \cdot (\alpha_0 \cdot P(S_0) + \alpha_1 \cdot P(S_1)) \tag{6}$$

$$= \text{wpr} \cdot \left(\alpha_0 \cdot \frac{1}{1 + \nu_1} + \alpha_1 \cdot \frac{\nu_1}{1 + \nu_1}\right) \tag{7}$$

$$= \text{wpr} \cdot \frac{\alpha_0 + \alpha_1 \cdot \nu_1}{1 + \nu_1} \tag{8}$$

(9)

Propensity of transcription when $N = 2$

When $N = 2$, there are four states S_j. The relative activations α_j can be defined such that H_1 and H_2 are independent (additive) or synergistic (multiplicative). In order to define the propensity of transcription $f(.)$, the Hill equation $P(S_j)$ is extended for two transcription factors.

Let w_j be the numerator of $P(S_j)$, defined as the product of all transcription factors bound in that state:

$$w_0 = 1 \tag{10}$$

$$w_1 = \nu_1 \tag{11}$$

$$w_2 = \nu_2 \tag{12}$$

$$w_3 = \nu_1 \cdot \nu_2 \tag{13}$$

The denominator of $P(S_j)$ is then equal to the sum of all w_j. The probability of state S_j is thus defined as:

$$P(S_j) = \frac{w_j}{\sum_{j=0}^{j<2^N} w_j} \tag{14}$$

$$= \frac{w_j}{1 + \nu_1 + \nu_2 + \nu_1 \cdot \nu_2} \tag{15}$$

$$= \frac{w_j}{\prod_{i=1}^{i\leq N}(\nu_i + 1)} \tag{16}$$

Substituting $P(S_j)$ and w_j into $f(.)$ results in the following equation:

$$f(y_1, y_2) = \text{wpr} \cdot \sum_{j=0}^{2^N-1} \alpha_j \cdot P(S_j) \tag{17}$$

$$= \text{wpr} \cdot \frac{\sum_{j=0}^{2^N-1} \alpha_j \cdot w_j}{\prod_{i=1}^{i\leq N}(\nu_i + 1)} \tag{18}$$

$$= \text{wpr} \cdot \frac{\alpha_0 + \alpha_1 \cdot \nu_1 + \alpha_2 \cdot \nu_2 + \alpha_3 \cdot \nu_1 \cdot \nu_2}{(\nu_1 + 1) \cdot (\nu_2 + 1)} \tag{19}$$

$$\tag{20}$$

Propensity of transcription for an arbitrary N

For an arbitrary N, there are 2^N states S_j. The relative activations α_j can be defined such that H_1 and H_2 are independent (additive) or synergistic (multiplicative). In order to define the propensity of transcription $f(.)$, the Hill equation $P(S_j)$ is extended for N transcription factors.

Let w_j be the numerator of $P(S_j)$, defined as the product of all transcription factors bound in that state:

$$w_j = \prod_{i=1}^{i\leq N} (j \bmod i) = 1 \; ? \; \nu_i : 1 \tag{21}$$

The denominator of $P(S_j)$ is then equal to the sum of all w_j. The probability of state S_j is thus defined as:

$$P(S_j) = \frac{w_j}{\sum_{j=0}^{j<2^N} w_j} \tag{22}$$

$$= \frac{w_j}{\prod_{i=1}^{i\leq N}(\nu_i + 1)} \tag{23}$$

Substituting $P(S_j)$ into $f(.)$ yields:

$$f(y_1, y_2, \ldots, y_N) = \text{wpr} \cdot \sum_{j=0}^{2^N-1} \alpha_j \cdot P(S_j) \tag{24}$$

$$= \text{wpr} \cdot \frac{\sum_{j=0}^{2^N-1} \alpha_j \cdot w_j}{\prod_{i=1}^{i\leq N}(\nu_i + 1)} \tag{25}$$

Propensity of transcription for a large N

For large values of N, computing $f(.)$ is practically infeasible as it requires performing 2^N summations. In order to greatly simplify $f(.)$, α_j could be defined as 0 when one of the regulators inhibits transcription and 1 otherwise.

$$\alpha_j = \begin{cases} 0 & \text{if } \exists i : j \bmod i = 1 \text{ and } H_i \text{ represses } G \\ 1 & \text{otherwise} \end{cases} \tag{26}$$

Substituting equation 26 into equation 25 and defining $R = \{1, 2, \dots, N\}$ and $R^+ = \{i | H_i \text{ activates } G\}$ yields the simplified propensity function:

$$f(y_1, y_2, \dots, y_N) = \text{wpr} \cdot \frac{\prod_{i \in R^+}(\nu_i + 1)}{\prod_{i \in R}(\nu_i + 1)} \tag{27}$$

Independence, synergism and basal expression

The definition of α_j as in equation 26 presents two main limitations. Firstly, since $\alpha_0 = 1$, it is impossible to tweak the propensity of transcription when no transcription factors are bound. Secondly, it is not possible to tweak the independence and synergism of multiple regulators.

Let $\text{ba} \in [0, 1]$ denote the basal expression strength G (i.e. how much will G be expressed when no transcription factors are bound), and $\text{sy} \in [0, 1]$ denote the synergism of regulators H_i of G, the transcription propensity becomes:

$$f(y_1, y_2, \dots, y_N) = \text{wpr} \cdot \frac{\text{ba} - \text{sy}^{|R^+|} + \prod_{i \in R^+}(\nu_i + \text{sy})}{\prod_{i \in R}(\nu_i + 1)} \tag{28}$$

Compute average expression along backbone transitions

When simulating the developmental backbone, we go through the edges of the backbone state network defined in an earlier step (Section), starting from the root state. It is assumed the root state has no modules active and has no expression of any molecules. To get to the next state, we follow a transition starting from the root state, activate and deactivate the modules as indicated by the transition, and compute the average molecule abundance along the transition. To compute the average abundance, we perform small time steps $t = 0.001$ and let each reaction (Section) occur t times its propensity.

Simulate single cells

dyngen uses Gillespie's stochastic simulation algorithm (SSA)[13] to simulate dynamic processes. An SSA simulation is an iterative process where at each iteration one reaction is triggered.

Each reaction consists of its propensity – a formula to calculate the probability of the reaction occurring during an infinitesimal time interval – and the effect – how it will affect the current state if triggered. Each time a reaction is triggered, the simulation time is incremented by $\tau = \frac{1}{\sum_j prop_j} \ln\left(\frac{1}{r}\right)$, with $r \in U(0, 1)$ and $prop_j$ the propensity value of the jth reaction for the current state of the simulation.

GillespieSSA2 is an optimised library for performing SSA simulations. The propensity functions are compiled to C++ and SSA approximations can be used which allow to trigger many reactions simultaneously at each iteration. The framework also allows to store the abundance levels of molecules only after a specific interval has passed since the previous census. By setting the census interval to 0, the whole simulation's trajectory is retained but many of these time points will contain very similar information. In addition to the abundance levels, also the propensity values and the number of firings of each of the reactions at each of the time steps can be retained, as well as specific sub-calculations of the propensity values, such as the regulator activity level $reg_{G,H}$.

Map SSA simulations to backbone

We compute the Pearson correlation between the state vectors in the simulation and the average expression levels along a transition in the backbone. Each timepoint in the SSA simulation is mapped to the point in the backbone that has the highest correlation value.

Simulate experiment

From the SSA simulation we obtain the abundance levels of all the molecules at every state. We need to replicate technical effects introduced by experimental protocols in order to obtain data that is similar to real data. For this, the cells are sampled from the simulations and molecules are sampled for each of the cells. Real datasets are used in order to achieve similar data characteristics.

Sample cells

In this step, N cells are sampled the simulations. Two approaches are implemented: sampling from an unsynchronised population of single cells (snapshot) or sampling at multiple time points in a synchronised population (time series).

Snapshot The backbone consists of several states linked together by transition edges with length L_i, to which the different states in the different simulations have been mapped (Figure 10A). From each transition, $N_i = N/\frac{L_i}{\sum L_i}$ cells are sampled uniformly, rounded such that $\sum N_i = N$.

Time series Assuming that the final time of the simulations is T, the interval $[0, T]$ is divided into k equal intervals of width w separated by $k-1$ gaps of width g. $N_i = N/k$ cells are sampled uniformly from each interval (Figure 10B), rounded such that $\sum N_i = N$. By default, $k = 8$ and $g = 0.75$. For usual dyngen simulations, $10 \leq T \leq 20$. For larger values of T, k and g should be increased accordingly.

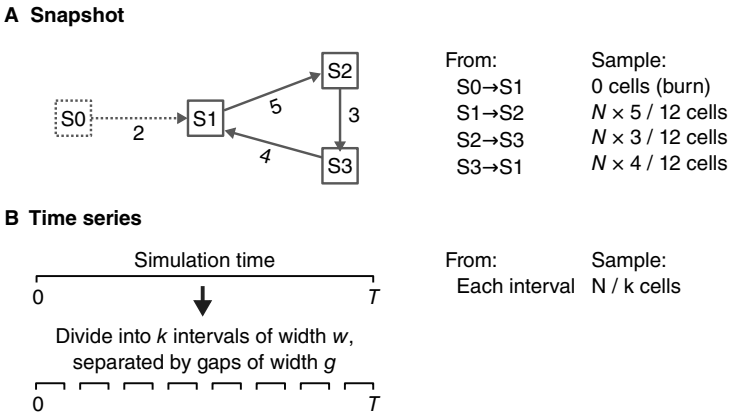

Figure 10: **Two approaches can be used to sample cells from simulations: snapshot and time-series.**

Sample molecules

Molecules are sampled from the simulation to replicate how molecules are experimentally sampled. A real dataset is downloaded from a repository of single-cell RNA-seq datasets[31]. For each *in silico* cell i, draw its library size ls_i from the distribution of transcript counts per cell in the real dataset. The capture rate cr_j of each *in silico* molecule type j is drawn from $N(1, 0.05)$. Finally, for each cell i, draw ls_i molecules from the multinomial distribution with probabilities $cr_j \times ab_{i,j}$ with $ab_{i,j}$ the molecule abundance level of molecule j in cell i.

Determining the casewise ground-truth regulatory network

Calculating the regulatory effect of a regulator R on a target T (Figure 6F) requires determining the contribution of R in the propensity function of the transcription of T (section) with respect to other regulators. This information is

useful, amongst others, for benchmarking casewise network inference methods.

The regulatory effect of R on T at a particular state S is defined as the change in the propensity of transcription when R is set to zero, scaled by the inverse of the pre-mRNA production rate of T. More formally:

$$\text{regeffect}_G = \frac{\text{proptrans}_G(S) - \text{proptrans}_G(S[y_T \leftarrow 0])}{\text{wpr}_G}$$

Determining the regulatory effect for all interactions and cells in the dataset yields the complete casewise ground-truth GRN (Figure 11). The regulatory effect lie between $[-1, 1]$, where -1 represents complete inhibition of T by R, 1 represents maximal activation of T by R, and 0 represents inactivity of the regulatory interaction between R and T.

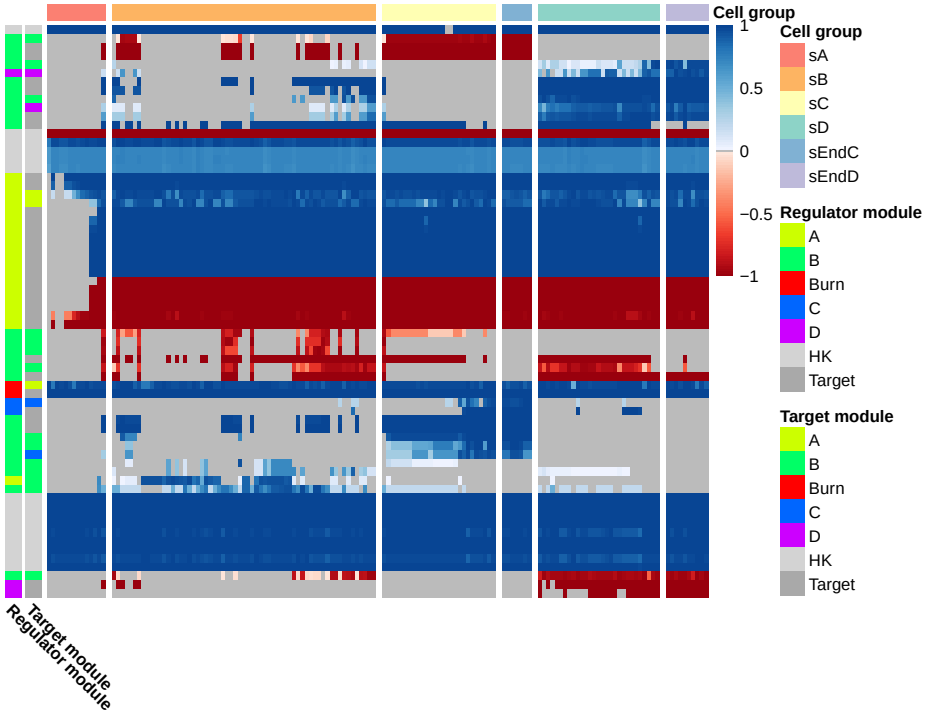

Figure 11: **The casewise regulatory effects of all interactions, computed on cells part of a bifurcation trajectory.** Negative values correspond to inhibitory interactions, positive values to activating interactions, and zero values correspond to inactive interactions.

Comparison of casewise network inference methods

Several datasets were generated using the different predefined backbones. For every cell in the dataset, the transcriptomics profile and the corresponding casewise ground-truth regulatory network was determined (Section).

Several casewise NI methods were considered for comparison: SCENIC[14], LIONESS[32, 15], and SSN[16].

LIONESS[@Kuijjer et al. [15]; kuijjer_lionessrsinglesample_2019] uses the Pearson correlation to infer casewise GRNs. To do so, first the Pearson correlation is calculated between regulators and targets for all samples. Next, the Pearson correlation is again calculated for all samples excluding one sample. The difference between the two correlation matrices is considered a casewise GRN for that particular profile. This process is repeated for all profiles, resulting in a casewise GRN.

SSN[16] has, in essence, the exact same methodology as LIONESS. It is worth noting that the LIONESS preprint was released before the publication of SSN. Since no implementation was provided by the authors, we implemented SSN in R using basic R and tidyverse functions[33] and marked results from this implementation as "SSN*".

SCENIC[14] is a pipeline that consists of four main steps. Step 1: classical network inference is performed with arboreto, which is similar to GENIE3[34]. Step 2: select the top 10 regulators per target. Interactions are grouped

together in 'modules'; each module contains one regulator and all of its targets. Step 3: filter the modules using motif analysis. Step 4: for each cell, determine an activity score of each module using AUCell. As a post-processing of this output, all modules and the corresponding activity scores are combined back into a casewise GRN consisting of (cell, regulator, target, score) pairs. For this analysis, the Python implementation of SCENIC was used, namely pySCENIC. Since dyngen does not generate motif data, step 3 in this analysis is skipped.

The AUROC and AUPR metrics are common metrics for evaluating a predicted GRN with a ground-truth GRN. To compare a predicted casewise GRN with the ground-truth casewise GRN, the top 10'000 interactions per cell were retained. For each cell-specific network, the AUROC and AUPR were calculated. ## Comparison of RNA velocity methods {#sec:dyngen-velcompare}

15 datasets were generated with 5 different backbones: linear, linear simple, bifurcating, cyclic, and disconnected. We extracted a ground truth RNA velocity by subtracting for each mRNA molecule the propensity of its production by the propensity of its degradation. If the expression of an mRNA will increase in the future, this value is positive, while it is negative if it is going to decrease. For each gene, we compared the ground truth velocity with the observed velocity by calculating the Spearman rank correlation.

We compared two RNA velocity methods. The velocyto method[18], as implemented in the velocyto.py package, in which we varied the "assumption" parameter between "constant_unspliced" and "constant_velocity". The scvelo method[19], as implemented in the python scvelo package (http://scvelo.de), in which we varied the "mode" parameter between "deterministic", "stochastic" and "dynamical". For both methods, we used the same normalized data as provided by dyngen, with no extra cell or feature filtering. We also matched the parameters between both methods as best as possible, i.e. the k parameter for smoothing was set to 20 for both methods.

To visualize the velocity on an embedding, we used the "velocity_embedding" function, implemented in the scvelo python package.

References

[1] Luke Zappia, Belinda Phipson, and Alicia Oshlack. "Splatter: Simulation of Single-Cell RNA Sequencing Data". In: *Genome Biology* 18 (Sept. 2017), p. 174. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1305-0.

[2] Beate Vieth et al. "powsimR: Power Analysis for Bulk and Single Cell RNA-Seq Experiments". In: *Bioinformatics* 33.21 (Nov. 1, 2017), pp. 3486–3488. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx435.

[3] Nikolaos Papadopoulos, Rodrigo Gonzalo Parra, and Johannes Soeding. "PROSSTT: Probabilistic Simulation of Single-Cell RNA-Seq Data for Complex Differentiation Processes". In: *bioRxiv* (Jan. 2018), p. 256941. DOI: 10.1101/256941.

[4] Xiuwei Zhang, Chenling Xu, and Nir Yosef. "Simulating Multiple Faceted Variability in Single Cell RNA Sequencing". In: *Nature Communications* 10.1 (June 13, 2019), pp. 1–16. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10500-w.

[5] Kelly Street et al. "Slingshot: Cell Lineage and Pseudotime Inference for Single-Cell Transcriptomics". In: *BMC Genomics* 19.1 (June 2018), p. 477. ISSN: 1471-2164. DOI: 10.1186/s12864-018-4772-0.

[6] R Gonzalo Parra et al. "Reconstructing Complex Lineage Trees from scRNA-Seq Data Using MERLoT". In: *bioRxiv* (Feb. 2018), p. 261768. DOI: 10.1101/261768.

[7] Edroaldo Lummertz da Rocha et al. "Reconstruction of Complex Single-Cell Trajectories Using CellRouter". In: *Nature Communications* 9.1 (Mar. 1, 2018), p. 892. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03214-y.

[8] Yingxin Lin et al. "scClassify: Hierarchical Classification of Cells". In: *bioRxiv* (Jan. 1, 2019), p. 776948. DOI: 10.1101/776948.

[9] Angelo Duò, Mark D. Robinson, and Charlotte Soneson. "A Systematic Performance Evaluation of Clustering Methods for Single-Cell RNA-Seq Data". In: *F1000Research* 7 (2018), p. 1141. ISSN: 2046-1402. DOI: 10.12688/f1000research.15666.2. pmid: 30271584.

[10] Wouter Saelens et al. "A Comparison of Single-Cell Trajectory Inference Methods". In: *Nature Biotechnology* 37 (May 2019). ISSN: 15461696. DOI: 10.1038/s41587-019-0071-9.

[11] Charlotte Soneson and Mark D. Robinson. "Bias, Robustness and Scalability in Single-Cell Differential Expression Analysis". In: *Nature Methods* 15.4 (Apr. 2018), pp. 255–261. ISSN: 1548-7105. DOI: 10.1038/nmeth.4612. pmid: 29481549.

[12] Tim Stuart and Rahul Satija. "Integrative Single-Cell Analysis". In: *Nature Reviews Genetics* 20.5 (May 2019), pp. 257–272. ISSN: 1471-0064. DOI: 10.1038/s41576-019-0093-7.

[13] Daniel T. Gillespie. "Exact Stochastic Simulation of Coupled Chemical Reactions". In: *The Journal of Physical Chemistry* 81.25 (Dec. 1, 1977), pp. 2340–2361. ISSN: 0022-3654. DOI: 10.1021/j100540a008.

[14] Sara Aibar et al. "SCENIC: Single-Cell Regulatory Network Inference and Clustering". In: *Nature Methods* (Oct. 2017). ISSN: 1548-7091. DOI: 10.1038/nmeth.4463.

[15] Marieke Lydia Kuijjer et al. "Estimating Sample-Specific Regulatory Networks". In: *iScience* 14 (Mar. 28, 2019), pp. 226–240. ISSN: 2589-0042. DOI: 10.1016/j.isci.2019.03.021. pmid: 30981959.

[16] Xiaoping Liu et al. "Personalized Characterization of Diseases Using Sample-Specific Networks". In: *Nucleic Acids Research* 44.22 (2016), e164–e164. ISSN: 0305-1048. DOI: 10.1093/nar/gkw772. pmid: 27596597.

[17] Amit Zeisel et al. "Coupled Pre-mRNA and mRNA Dynamics Unveil Operational Strategies Underlying Transcriptional Responses to Stimuli". In: *Molecular Systems Biology* 7.1 (Jan. 1, 2011), p. 529. ISSN: 1744-4292. DOI: 10.1038/msb.2011.62.

[18] Gioele La Manno et al. "RNA Velocity of Single Cells". In: *Nature* 560.7719 (Aug. 2018), pp. 494–498. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0414-6.

[19] Volker Bergen et al. "Generalizing RNA Velocity to Transient Cell States through Dynamical Modeling". In: *bioRxiv* (Oct. 29, 2019), p. 820936. DOI: 10.1101/820936.

[20] Thomas Schaffter, Daniel Marbach, and Dario Floreano. "GeneNetWeaver: In Silico Benchmark Generation and Performance Profiling of Network Inference Methods." In: *Bioinformatics* 27.16 (Aug. 2011), pp. 2263–2270. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btr373. pmid: 21697125.

[21] Adam D. Ewing et al. "Combining Tumor Genome Simulation with Crowdsourcing to Benchmark Somatic Single-Nucleotide-Variant Detection". In: *Nature Methods* 12.7 (July 2015), pp. 623–630. ISSN: 1548-7105. DOI: 10.1038/nmeth.3407.

[22] Stephen Smith and Ramon Grima. "Spatial Stochastic Intracellular Kinetics: A Review of Modelling Approaches". In: *Bulletin of Mathematical Biology* 81.8 (Aug. 1, 2019), pp. 2960–3009. ISSN: 1522-9602. DOI: 10.1007/s11538-018-0443-1.

[23] N. Rekhtman et al. "Direct Interaction of Hematopoietic Transcription Factors PU.1 and GATA-1: Functional Antagonism in Erythroid Cells". In: *Genes & Development* 13.11 (June 1, 1999), pp. 1398–1411. ISSN: 0890-9369. DOI: 10.1101/gad.13.11.1398. pmid: 10364157.

[24] Heping Xu et al. "Regulation of Bifurcating {B} Cell Trajectories by Mutual Antagonism between Transcription Factors {IRF4} and {IRF8}". In: *Nat. Immunol.* 16.12 (Dec. 2015), pp. 1274–1281.

[25] Thomas Graf and Tariq Enver. "Forcing Cells to Change Lineages". In: *Nature* 462.7273 (Dec. 2009), p. 587. ISSN: 1476-4687. DOI: 10.1038/nature08533.

[26] Jin Wang et al. "Quantifying the Waddington Landscape and Biological Paths for Development and Differentiation". In: *Proceedings of the National Academy of Sciences* 108.20 (May 2011), pp. 8257–8262. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1017017108. pmid: 21536909.

[27] James E Ferrell. "Bistability, Bifurcations, and Waddington's Epigenetic Landscape". In: *Current Biology* 22.11 (June 2012), R458–R466. ISSN: 0960-9822. DOI: 10.1016/j.cub.2012.03.045.

[28] Nir Yosef et al. "Dynamic Regulatory Network Controlling {TH17} Cell Differentiation". In: *Nature* 496.7446 (2013), pp. 461–468.

[29] Marina Lizio et al. "Gateways to the FANTOM5 Promoter Level Mammalian Expression Atlas". In: *Genome Biology* 16.1 (Jan. 5, 2015), p. 22. ISSN: 1465-6906. DOI: 10.1186/s13059-014-0560-6.

[30] Michael A. Gibson and Jehoshua Bruck. "A Probabilistic Model of a Prokaryotic Gene and Its Regulation". In: *Computational Methods in Molecular Biology: From Genotype to Phenotype, MIT press, Cambridge* (2000).

[31] Robrecht Cannoodt et al. "Single-Cell -Omics Datasets Containing a Trajectory". In: *Zenodo* (Oct. 2018). DOI: 10.5281/zenodo.1211532.

[32] Marieke Lydia Kuijjer et al. "Estimating Sample-Specific Regulatory Networks". In: (2015), pp. 1–19. URL: http://arxiv.org/abs/1505.06440.

[33] Hadley Wickham et al. "Welcome to the Tidyverse". In: (Nov. 21, 2019). DOI: 10.21105/joss.01686.

[34] Vân Anh Huynh-Thu et al. "Inferring Regulatory Networks from Expression Data Using Tree-Based Methods". In: *PLoS ONE* 5.9 (Jan. 2010), e12776. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0012776. pmid: 20927193.