



Modeling Progression of Single Cell Populations Through the Cell Cycle as a Sequence of Switches

Andrei Zinovyev^{1,2,3*}, Michail Sadovsky^{4,5,6,7}, Laurence Calzone^{1,2,3}, Aziz Fouché^{1,2,3}, Clarice S. Groeneveld^{8,9}, Alexander Chervov^{1,2,3}, Emmanuel Barillot^{1,2,3} and Alexander N. Gorban^{7,10}

¹Institut Curie, PSL Research University, Paris, France, ²INSERM, Paris, France, ³MINES ParisTech, PSL Research University, C BIO-Centre for Computational Biology, Paris, France, ⁴Institute of Computational Modeling (RAS), Krasnoyarsk, Russia, ⁵Laboratory of Medical Cybernetics, V.F.Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, Russia, ⁶Federal Research and Clinic Center of FMBA of Russia, Krasnoyarsk, Russia, ⁷Laboratory of Advanced Methods for High-Dimensional Data Analysis, Lobachevsky University, Nizhniy Novgorod, Russia, ⁸Cartes d'Identité des Tumeurs (CIT) Program, Ligue Nationale Contre le Cancer, Paris, France, ⁹Oncologie Moléculaire, UMR144, Institut Curie, Paris, France, ¹⁰Department of Mathematics, University of Leicester, Leicester, United Kingdom

OPEN ACCESS

Edited by:

Alberto Jesus Martin,
Universidad Mayor, Chile

Reviewed by:

Anand Banerjee,
Virginia Tech, United States
Paola Lecca,
Free University of Bozen-Bolzano, Italy

*Correspondence:

Andrei Zinovyev
andrei.zinovyev@curie.fr

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 12 October 2021

Accepted: 15 December 2021

Published: 01 February 2022

Citation:

Zinovyev A, Sadovsky M, Calzone L,
Fouché A, Groeneveld CS, Chervov A,
Barillot E and Gorban AN (2022)
Modeling Progression of Single Cell
Populations Through the Cell Cycle as
a Sequence of Switches.
Front. Mol. Biosci. 8:793912.
doi: 10.3389/fmolb.2021.793912

Cell cycle is a biological process underlying the existence and propagation of life in time and space. It has been an object for mathematical modeling for long, with several alternative mechanistic modeling principles suggested, describing in more or less details the known molecular mechanisms. Recently, cell cycle has been investigated at single cell level in snapshots of unsynchronized cell populations, exploiting the new methods for transcriptomic and proteomic molecular profiling. This raises a need for simplified semi-phenomenological cell cycle models, in order to formalize the processes underlying the cell cycle, at a higher abstracted level. Here we suggest a modeling framework, recapitulating the most important properties of the cell cycle as a limit trajectory of a dynamical process characterized by several internal states with switches between them. In the simplest form, this leads to a limit cycle trajectory, composed by linear segments in logarithmic coordinates describing some extensive (depending on system size) cell properties. We prove a theorem connecting the effective embedding dimensionality of the cell cycle trajectory with the number of its linear segments. We also develop a simplified kinetic model with piecewise-constant kinetic rates describing the dynamics of lumps of genes involved in S-phase and G2/M phases. We show how the developed cell cycle models can be applied to analyze the available single cell datasets and simulate certain properties of the observed cell cycle trajectories. Based on our model, we can predict with good accuracy the cell line doubling time from the length of cell cycle trajectory.

Keywords: cell cycle, mathematical modeling, molecular switches, transcription epoch, single cell data

1 INTRODUCTION

Progression through the cell cycle represents a complex dynamical process, regulated at multiple levels including transcriptome and proteome. The major components of it have been characterized (Hunt, 1991; Hunt et al., 2011), and a complex molecular machinery has been revealed (Tyson, 1991). Nevertheless, many aspects of cell cycle functioning remain to be elucidated (Giotti et al., 2019).

Progression through the cell cycle can be seen as a trajectory in multidimensional space of all possible cellular states, similar to other processes such as differentiation or ageing. However, this trajectory is characterized by special properties because it represents a periodic process. From an oversimplified perspective, at the end of this trajectory, a cell splits into two daughter cells twice as small, where each daughter cell has a state identical to the initial state of its parent. This requirement imposes certain constraints on the geometry and underlying mechanisms of the cell cycle trajectory (CCT), which could be reproduced with a mathematical model.

Cell cycle process has been a subject of mathematical modeling for many decades (Chen et al., 2004a; Ingolia and Murray, 2004; Sible and Tyson, 2007). Most of the existing models focused on reproducing the regulatory logics at the level of protein expression, protein-protein interactions and post-translational modifications. Multiple modeling formalisms have been used such as chemical kinetics (Tyson, 1991; Chen et al., 2004b), logical modeling (Fauré et al., 2006; Deritei et al., 2019), Petri nets (Kotani, 2002), or approaches based on tropical algebra (Noel et al., 2012; Radulescu et al., 2012). A hybrid approach, combining discrete, governed by Boolean dynamics, and continuous, governed by chemical kinetics, variables was suggested to model cell cycle (Singhania et al., 2011; Noël et al., 2013). The mathematical description of the cell cycle transcriptional dynamics has not yet been thoroughly addressed, though.

High-throughput omics measurements gave rise to a number of molecular studies with the objective to characterize each cell cycle phase in terms of their associated molecular changes, i.e., sets of specifically expressed genes (Dominguez et al., 2016; Giotti et al., 2019). The appearance of single cell technologies reinforced the interest towards the description of the molecular organization of the cell cycle for several reasons. First, it allows the visualization of the cell cycle trajectory explicitly without synchronizing individual cells, which can be problematic, especially *in vivo*. Then, recent single cell transcriptomic and proteomic studies provide molecular description of progression through the cell cycle in a continuous fashion. Such representation attempts to delineate the cell cycle phase borders and also characterizes each cell for its precise progression position within each phase (Leng et al., 2015; Liu et al., 2017; Hsiao et al., 2020; Mahdessian et al., 2021).

A thorough understanding of cell cycle functioning is of utmost importance for cancer research, where the deviation from the normal cell cycle progression is expected. A number of questions can be raised: What is the normal pattern of the events comprising a cell cycle, and to what extent does it vary in normal physiology? What deviations from a normal cell cycle are characteristic for a tumor cell? What processes trigger these changes and are they specific to a cancer type? and many others.

Some mathematical models of the cell cycle try to tackle these questions. For example, agent-based or cellular automaton cell cycle models focus on the optimization of cancer drug delivery (Altinok et al., 2007), competition of fast and slow cell cycles within a tumor under treatment (Tzamali et al., 2020), or cell confluence and elongation of the G1 phase (Bernard et al., 2019). However, most of the existing models remain limited to describe the behavior of cell

cycle during tumorigenesis at full complexity because of the existing discrepancy between the nature of the available molecular data and the level of the details of these models. Thus, the most comprehensive data source currently available is at the level of transcriptomic changes in single cells, while the existing modeling efforts focus on protein players. The data reveal the role of hundreds of genes and proteins in cell cycle dynamics, while the models include a tiny fraction of this number. Therefore, we believe that the development of mathematical models matching the scale and the nature of the abundant available data is still highly needed. In particular, even a simple mechanistic model of cell cycle transcriptome dynamics, capturing its main features, is lacking in the field. It appears that using dynamical variables representing relatively large lumps of genes (e.g., all genes involved in DNA replication) might be a useful coarse-grained approach to model cellular transcriptomes, which is one motivation of this study.

Single cell studies of cell cycle trajectories in snapshots of actively proliferating cells represent a unique opportunity to formulate the most general principles of cell cycle functioning. A recent study has introduced the principle of minimizing transcriptomic acceleration (Schwabe et al., 2020), which suggests that the transcriptomic cell cycle trajectory represents a spiral, or, after neglecting the relatively slow drift unrelated to cell cycle progression, a shape close to “a flat circle”. This type of trajectories was indeed phenomenologically observed in the HeLa cell line profiled with scRNASeq technology, after deconvoluting the transcriptomic dynamics connected to the cell cycle from other sources of transcriptional heterogeneity. In particular, the absence of cell cycle-related transcriptional epochs was deduced from this model.

In the current study, we suggest an extended and different point of view on the properties of transcriptomic cell cycle trajectory which, in our opinion, in some cases better matches its observed properties in various cellular systems, when sufficiently good quality data can be collected. We propose a formal model of CCT as a sequence of epochs of growth during each of which the trajectory is approximately linear in the space of logarithmic coordinates. Therefore, CCT can be modeled as a piecewise linear trajectory in the space of logarithms of some extensive cell properties, followed by a shift at the vector with coordinates $-\log 2$ which represents the cell division event. This model explicitly assumes existence of well-defined transcriptional epochs in CCT.

Movement along linear trajectory in the space of logarithms of the values of some cellular properties means that along the trajectory any two such properties x_i, x_j are connected through a power law dependence $x_i = \alpha x_j^\beta$, $\alpha, \beta = \text{const}$. Such dependencies are known as allometric in many fields of biology (Holford and Anderson, 2017; Packard, 2017; White et al., 2019; Pretzsch, 2020; Zhou et al., 2021). Some approaches in mathematical chemistry and theoretical biology, dealing with systems in stable non-equilibrium, exploit the linear relations between chemical potentials which can be expressed as logarithms of species concentrations (Bauer, 1935; Gorban, 2018).

Particular cases of allometric dependencies are when all the quantities grow linearly with physical time, or when all the quantities follow exponential growth or decay $x_i = b_i \exp(a_i t)$. The model of movement along piecewise linear trajectories with an event of cell division represents the simplest scenario, easy to

simulate and analyse theoretically. Nevertheless, the most important conclusions derived from this analysis will stay valid for the trajectories that do not deviate too much from linearity.

Using the model of piecewise linear growth with division, we formulate a fundamental statement about correspondence between the number of linear segments in the cell cycle trajectory m , which corresponds to a number of the most important states of the cell cycle-related transcriptional machinery, and its effective embedding dimension n . The first part of the statement, $m \geq n$, can be described as a strict theorem with formal proof, whereas the second part, $m \leq n$, can be formulated as a feasible hypothesis, that can be validated using available data. The correspondence $m = n$ suggests that the embedding dimensionality of the transcriptomic cell cycle trajectory is larger than 2, since the number of segments we can observe can be as high as four or five. This allows us to state that the shape of the cell cycle trajectory is essentially not flat.

The type of models discussed here was partly introduced by *Shkolnik* (a pseudonym for a collective authorship), including authors of this manuscript (Shkolnik, 1989). Here, we significantly extend the previous effort and adapt it to the description of the cell cycle trajectory in single cell datasets.

In order to connect the geometric properties of cell cycle trajectory to interpretable mechanistic parameters, we extended the model of piecewise linear growth in logarithmic coordinates, to a simple kinetic model with rates depending on time as piecewise constant functions. In this case, some of the segments of the trajectory become nonlinear but remain smooth and do not deviate from linearity too far. Therefore, the suggested model is conceptually similar to previously suggested hybrid discrete-continuous models, but conceptualizes them, addresses the transcriptional dynamics and can be fit to multiple available scRNASeq datasets (Singhania et al., 2011; Noël et al., 2013).

The suggested cell cycle modeling framework and the representation of the cell cycle progression as a system of switches allows us to 1) determine which genes play the most important role in each transcriptional epoch, in a concrete system under study, 2) compare the genes related to the same transcriptional epoch between two biological systems or conditions, 3) predict the ratios between physical time durations of the transcriptional epochs, 4) predict the effect of shortening of certain transcriptional epochs on the shape of the cell cycle trajectory and transcriptional dynamics of the related groups of genes, and 5) predict the doubling time of proliferating cell populations from the length of the cell cycle trajectory observed in single cell scRNASeq data. The suggested framework can be exploited to study the cell cycle in various systems, from cell lines to tumors.

2 METHODS AND MATERIALS

2.1 Single Cell Data Used in This Study

We made a screening of available single cell sequencing of cancer cell lines in order to identify datasets with sufficient number of good quality single cell transcriptomic profiles and in which the principal source of transcriptomic heterogeneity would be

progression through the cell cycle. We identified publicly available scRNASeq data on CHLA9 Ewing sarcoma cell line, produced with 10x Genomics sequencing technology (Miller et al., 2020), which contained more than 4,000 cells with total number of unique molecular identifiers (UMIs) varying from 10,000 to 50,000 per cell, after applying the standard quality criteria and filtering cells containing a large fraction (>20%) of reads in mitochondrial genes. For this dataset, we reanalyzed the raw sequencing data using Kallisto mapper (Bray et al., 2016) resulting in a loom file that could be used for obtaining the gene expression levels and for quantifying RNA velocity vectors (La Manno et al., 2018).

In addition, we used a recently published collection of 200 scRNASeq profiles of cancer cell lines from Cancer Cell Line Encyclopedia (CCLE) collection (Kinker et al., 2020). We also analyzed several scRNASeq datasets by downloading them directly from Gene Expression Omnibus (GEO).

The estimation of cell line doubling times, when available were obtained from Cellosaurus database (Bairoch, 2018).

2.2 Definition of Cell Cycle Genes

We systematically tested several existing definitions of cell cycle gene sets and verified that our results remain qualitatively invariant even if the choice of cell cycle gene set can vary. In our experiments, we used the following cell cycle gene set definitions:

- Standard “Regev’s set”: markers of S- and G2/M cell cycle phases used in scanpy tutorials (Tirosh et al., 2016)
- Set of cell cycle genes annotated in Reactome pathway database (Jassal et al., 2020)
- Set of top-contributing genes, extracted from application of independent component analysis (ICA) to the dataset under study, from those components whose top-contributing genes were strongly associated with the cell cycle. In particular, similar to our previous work (Aynaud et al., 2020), two independent components were significantly enriched with the markers of S- and G2/M cell cycle phases in all single cell cell line datasets we analyzed.

Cell cycle phase scores were computed as an average expression of marker genes for the corresponding cell cycle phase in log scale, which roughly corresponds to the geometric mean of the raw count measures.

2.3 Pooling Reads From Neighbouring Cells for Compensating the Technical Drop-Out

We found out that the cell cycle trajectories appear less noisy and more tractable by trajectory inference methods when standard pooling approach was applied to the raw count data, using an initial estimate of cell-to-cell proximity. More precisely, we used the initial standard data normalization and dimensionality reduction in order to compute the distances between cells and construct the initial kNN graph, which was used to pool row reads from a cell and all its k nearest neighbours. In our experiments, we used $k = 10$ and $n = 30$ components for reducing the data

dimensionality during normalization. Pooled read counts were used for final normalization, but the initial total read counts per cell measure were kept for visualization and further analysis.

2.4 Cell Cycle Trajectory-Based Single Cell Data Normalization

Total number of reads in a cell represents a strong signal in proliferating cell populations. By itself, it is an extensive value such that it should be divided (approximately) by half in the process of cell division. In our modeling approach, we needed a description of the cell state in terms of extensive values of gene expression levels measured such that they would be also divided approximately by two on average after the moment of cell division. Therefore, the widely used global library size normalization did not suit our purposes, since after global library size normalization, cell division does not lead to halving the total number of reads.

At the same time we observed that without any library size normalization, the cells presumably located at similar stages of cell cycle progression could be characterized by a wide range of total number of reads, probably caused by technical variability factors. Therefore, library size normalization was required but not at the global cell population level. We hypothesized that the total number of reads should increase in the course of cell cycle progression on average such that the cells characterized by similar value of pseudotime along the cell cycle trajectory could be normalized to the same local library size. As usual, this poses a chicken-or-egg problem because for reconstructing the cell cycle trajectory one needs normalized data, and for normalization of the library size one needs a reconstructed trajectory. This problem is similar to those approaches which use normalization locally conditioned on clusters in single cell datasets (Azizi et al., 2018).

We used a simplified two-stage approach for library size normalization which preserved both the geometric structure of CCT and the trend of increasing the total number of reads along CCT.

1. The row count data have been normalized to the global median number of counts and $\ln(x+1)$ -transformed, using standard functions of scanpy. 10,000 most variable genes have been selected, the dimensionality was reduced to 30 by PCA. In the reduced space, a kNN graph has been computed using the standard Euclidean distance for $k = 10$. This graph was used for pooling reads from neighbor cells as described above.
2. For such initially normalized dataset, we computed closed cell cycle trajectory in the subspace of cell cycle genes, by fitting a principal closed curve, using the Python implementation of ELPiGraph (Albergante et al., 2020). The data points were partitioned according to the proximity to the nodes of the elastic principal curve.
3. In each partition, we analyzed the distribution of the total number of reads across cells. We performed correction of cell-to-node assignment by splitting an anomalously wide partition between two neighboring partitions. The anomalously wide partition corresponded to the moment of cell division since it

contained both cells at the very end of cell cycle progression with the largest number of reads and cells just after cell division event containing the minimal number of reads. Splitting this distribution allowed us to distinguish cells just before and just after the cell division into distinct partitions.

4. The median total number of counts in each resulting corrected partition was computed. The median values of the total number of reads in the cells of each partition have been smoothed by univariate spline or a piecewise-linear function of pseudotime, taking into account the cyclic boundaries of the trajectory.
5. Each cell's library size was normalized to the smoothed local median value of the total number of reads.
6. The newly normalized pseudocount data matrix passed through the same pre-processing as described in 1), namely
 - a) Pooling reads from neighbour cells using the kNN graph obtained with trajectory-based normalized data,
 - b) $\ln(x+1)$ transformation, selecting most variable 10,000 genes.

The cell cycle trajectory-based normalization procedure is illustrated in the Jupyter notebook at https://github.com/auranic/CellCycleTrajectory_SegmentModel, which can be easily reused for other cell lines.

2.5 Computing the Cell Cycle Trajectory and Quantifying Pseudotime

We used the ELPiGraph Python package to fit elastic principal curves or closed elastic principal curves (principal circles) to single cell data distributions (Albergante et al., 2020). ELPiGraph was applied in the data space defined by the set of 10,000 most variable genes or by the cell cycle-related genes, after dimensionality reduction by PCA (first 30 principal components were retained). In order to compute open elastic principal curve with q nodes, first a closed curve was fit with $q/2$ nodes, then a node with the least number of data points projected onto it was removed from the principal graph, and this configuration was used as an initialization to compute the elastic principal graph without branching and having q nodes.

The pseudotime s_i for a data point x_i was computed as a continuous geodesic distance measured from the root node to the projection of x_i onto the principal curve, quantified in the units of the number of edges. Therefore, the value of the pseudotime was in the range $[0, q - 1]$, where q is the number of nodes. The root of the principal curve was chosen as one of its ends, such that the value of the initial total number of reads would increase as a function of pseudotime.

2.6 Curvature Analysis of the Cell Cycle Trajectory

In order to compute the Riemannian curvature of the principal curve defined by the position of its nodes in the multi-dimensional space $y_i \in R^n$, $i = 1, \dots, q$, the node coordinates were first represented as n functions of the natural parameter (pseudotime) s , $y_i^k = y_i^k(s_i)$, $i = 1, \dots, q$, $k = 1, \dots, n$. The value s_i for each node was taken as a number of edges of the elastic

principal curve connecting the node i to the root node. Each set of numbers $y_i^k(s_i), i = 1, \dots, q$ was interpolated by a cubic univariate spline $y^k(s)$. In each node i of the curve the curvature was evaluated as $R_i = \sum_{k=1}^n \left(\frac{d^2 y^k(s)}{ds^2} \Big|_{s=s_i} \right)^2$.

2.7 Estimating the Effective Dimensionality of a Set of Vectors

In order to estimate the effective dimensionality of CCT, we used scikit-dimension Python package (Bac et al., 2021). We used linear estimators of global intrinsic dimensionality, based on application of PCA and various approaches to select the significant number of eigenvalues from the scree plot.

In order to compute the effective rank of a rectangular matrix, we looked at the distribution of its singular values, and selected such a number of them that the ratio between the largest and the smallest number would not exceed 10, such that the reduced matrix is well-conditioned.

3 RESULTS

3.1 Example of a Cell Cycle Trajectory Extracted From Single Cell Data

The current study is motivated by the observation that after appropriate pre-processing of single cell RNA-Seq data (see Methods), one can observe the cell cycle trajectory (Figure 1) which can be approximated by a piecewise linear curve, with a gap between the beginning and the end of the trajectory corresponding to the cell division moment.

Here we use the example of Ewing sarcoma cell line CHLA9 sequenced at single cell level using the Chromium 10x technology (Miller et al., 2020). The distinguishing feature of this dataset was that it contained a significant number of proliferating cells with single cell transcriptomes of good quality (more than 4,000 cells with the total number of Unique Molecular Identifiers (UMIs) between 10,000 and 50,000). Also, the proliferation signal in this dataset seems to explain the largest fraction of transcriptomic heterogeneity, since in the plane of the first two principal components one can clearly observe the cyclic trajectory. In other cell line single cell datasets, the proliferative signal can be masked by other sources of transcriptomic heterogeneity, requiring special procedures of data treatment to reveal it (Aynaud et al., 2020; Liang et al., 2020; Schwabe et al., 2020).

The scRNA-Seq data have been normalized in order to preserve the pattern of dynamics of the total number of counts (UMIs) along the CCT, see Methods section. The normalized gene expression levels are represented at the logarithmic scale, following the standard practice. The multi-dimensional distribution of single cell transcriptomic profiles projected into the space of the first 30 principal components has been approximated by a principal curve (see Methods). The curvature of the principal curve has been estimated using the standard formulas of differential geometry, which revealed the existence of curvature peaks, and reflecting the rapid turning points of the trajectory. We hypothesized that these turning points correspond to the large-scale changes in the

transcriptional programs of the cell cycle process. The pattern of momentary velocities of the transcriptomic changes, estimated with RNA velocity, was compatible with this hypothesis (Figure 1A).

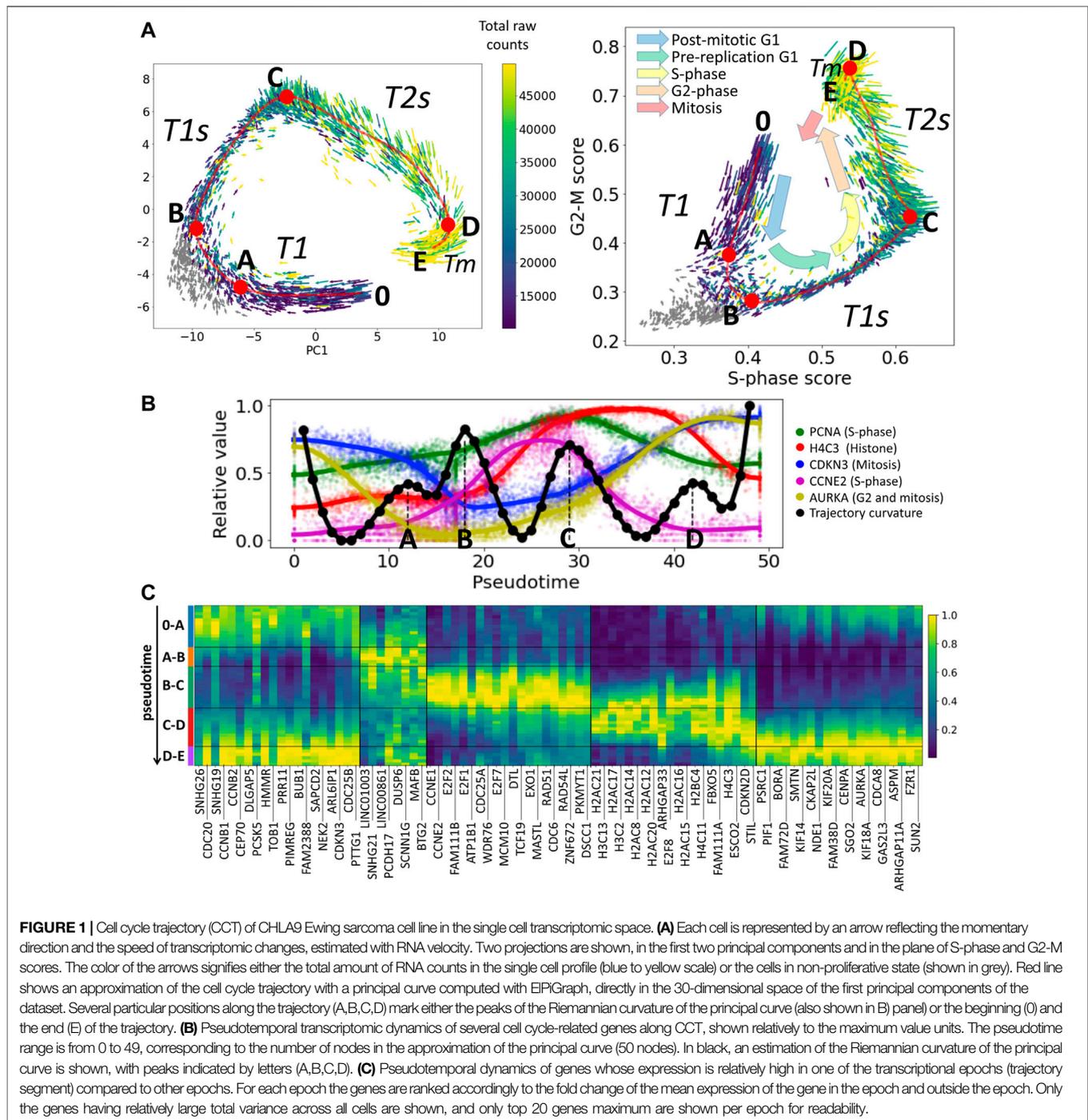
The pseudo-temporal dynamics of the known cell cycle-related genes confirmed that the trajectory curvature peaks delineate biologically meaningful transcriptional epochs. The epoch 0-A-B can be understood as an early G1 phase of the cell cycle, B-C as significantly overlapping with late G1 and S-phases, and C-D as overlapping with S- and G2-phases. The epoch D-E can presumably reflect the relatively short M phase (mitosis). Analysis of pseudotemporal gene expression dynamics inferred for this cell cycle trajectory shows that known cell cycle genes such as different cyclin types or E2F transcription factors have behaviour compatible with our interpretation (Figure 1C). We denote the identified transcriptional epochs as T1, T1s, T2s and Tm.

The switches between transcriptional epochs should not be confused with the action of cell cycle checkpoints that delineate cell cycle phases. The connection between the known molecular checkpoint mechanisms involving mainly protein-protein interactions and post-translational protein modifications and the transcriptional epochs might not be trivial or direct: partly, due to the delay between the gene and protein expression, and partly due to different parameters and constraints on the transcriptional and protein-protein interaction dynamics.

We can clearly observe the existence of the restriction point at the level of single cell transcriptome. In our notations, it belongs to the A-B segment of the cell cycle trajectory shown in Figure 1A, right. This transcriptional epoch separates post-mitotic (denoted as T1) and pre-replication parts of G1 phase, which corresponds to the classical definition of the R-point (e.g., from (Zetterberg et al., 1995)). Interestingly, in Figure 1A, right, one can observe that RNA velocity vectors reflect cells exiting from cell cycle and re-entering the cell cycle in the epoch between A and B turning points. Just after this transcriptional epoch, the expression of E2F transcription factors and Cyclin E start to increase as expected (Figure 1C).

We can also observe how, during each particular epoch, the components of a specific checkpoint mechanism are transcriptionally produced “just in time”. For example, components of the G1 DNA damage checkpoint (e.g., CDC25A, CDKN1A) are produced during the T1s epoch of the cell cycle trajectory where the S phase starts, the components of G2 DNA damage checkpoints (e.g., CDC25B, CDC25C, CHEK2) are produced in the late part of the C-D epoch (T2s), and spindle checkpoint components (e.g., CDC20) are transcriptionally abundant during the mitosis-related epoch D-E (Tm) and after the cell division in T1 (Figure 1C). In this sense, the transcriptional dynamics prepare the correct ground for a proper succession of post-transcriptional events but the exact borders of the transcriptional epochs do not have to match the precise checkpoint timing.

Remarkably, within each of the identified transcriptional cell cycle epochs, the global dynamics of the transcriptome remain close to linear in the logarithmic scale. This allows us to suggest a simple model which can, for example, represent the collective



dynamics of the genes related to the S-phase and G2/M phases (see below).

3.2 Model of Cell Cycle as a Trajectory of Allometric Growth With Switches and Divisions

Based on the observations of the properties of the cell cycle trajectory in several scRNASeq datasets, we hypothesized that it

can be recapitulated by a formal model of linear growth in logarithmic coordinates with switches and a cell division event. The suggested model is hybrid in nature, similar to some previously published models (Singhania et al., 2011; Noël et al., 2013). Namely, we distinguish the extrinsic observable cell state, characterized by continuous variables, and the intrinsic hidden cell state, characterized by discrete variables. The intrinsic state of a cell determines the parameters of the extrinsic dynamic process as in (Singhania et al., 2011).

Let the extrinsic state of a proliferating cell be determined by n substances quantified by their amounts, not their concentrations. Instead of their natural units (such as RNA counts), let us use the logarithms of these amounts. The cell is represented as an n -dimensional vector, and all possible combinations of these vector components define the cell configuration space. For our model, it is important that the considered n quantities are extensive measures, not intensive ones. Extensiveness here means that the total amount of a substance is a sum of the amounts found in different parts of a cell. A division (for two almost equal) daughter cells is formalized as a shift by the vector with all components equal $-\log 2$ in this space. A relevant example of extensive quantity is the total amount of RNA molecules present in a cell, or the amount of any specific subset of RNA molecules, i.e., representing mRNAs of the genes involved in a particular process (such as mitosis or S-phase).

We assume that there exists a finite discrete set of intrinsic cell states. In each of these states, the cell follows a linear trajectory in the extrinsic and continuous cell state space. This trajectory extends until the cell meets a condition, where a switch into another intrinsic state of the cell happens, which changes the direction of the trajectory. For simplicity, we assume that the conditions of a switch can be described by a linear function. The cell movement continues until a particular condition is met in which the cell division event is triggered leading to the aforementioned translation of the vector representing the extrinsic cell state.

Let us introduce some mathematical notations and consider a deterministic automaton A whose complete state is represented by a pair (x, s) , where $x \in R^n$ is a vector in n -dimensional continuous space (extrinsic state), and $s \in S$ is an integer number from a finite set $S = \{S_1, \dots, S_m\}$ (intrinsic state). In the rest of the study, we will call x a position of A and s an intrinsic state of A . We will denote the automaton A in position x and in the intrinsic state s as $A(x|s)$.

Each intrinsic state S_k is parameterized by a vector $a_k \in R^n$, $k = 1..m$ and by a linear manifold D_k of dimensionality $n - 1$ embedded in R^n (hyperplane), which we will call "the cell division hyperplane". D_k can be undefined, in this case, we denote $D_k = null$.

Let us also introduce a set of p functions $G = \{g_1, \dots, g_p\}$, $g_i: S \rightarrow S$, which we will call switches. Each switch g_i is a map which converts an intrinsic state $s_j \in S$ into another intrinsic state $s_r \in S$. Each switch g_i is parametrized by a hyperplane L_i existing in R^n and inducing the switch function g_i each time the trajectory of the automaton intersects L_i (see **Figure 2A**).

Finally, we introduce the cell division event ϕ which is a map between two states of A , such that $\phi((x, s)) \rightarrow (x + d, s_d)$, where $d \in R^n$ is a vector with negative components, and $s_d \in S$ is one of the possible intrinsic states of A .

We will characterize any hyperplane here by a linear functional $f(x|b, c) = b + \langle c, x \rangle$, $b \in R$, $c \in R^n$, where \langle, \rangle denotes the standard scalar product between two vectors. Using such a functional, for any pair of vectors $x_i, x_j \in R^n$ we can determine if the linear segment connecting x_i and x_j intersects the hyperplane or not. If the segment intersects the hyperplane then $f(x_i)f(x_j) < 0$, and if it does not intersect then $f(x_i)f(x_j) > 0$. $f(x_i)f(x_j) = 0$

is satisfied only in a non-general position when either x_i or x_j is located exactly on the hyperplane.

The update rules for the automaton A are described as follows. The automaton is in some initial position x_0 and the intrinsic state s_0 . It starts to move along the linear trajectory described by the equation $x = x_0 + a_0t$, where a_0 is the vector of movement associated with the state s_0 . This movement continues unless one of the two events happens. In the first case, A reaches the corresponding cell division plane D_0 (in case D_0 is not null). Then, the cell division event is triggered, $A(x|s) \rightarrow A(x + d|s_d)$. In the second case, x reaches a switch hyperplane L_j and then a switch of the intrinsic state of A happens without changing its position, $A(x|s_0) \rightarrow A(x|g_j(s_0))$. The movement continues along a new trajectory, corresponding to the new cell state, following the same rules: either the trajectory hits the cell division hyperplane or any of the switch planes.

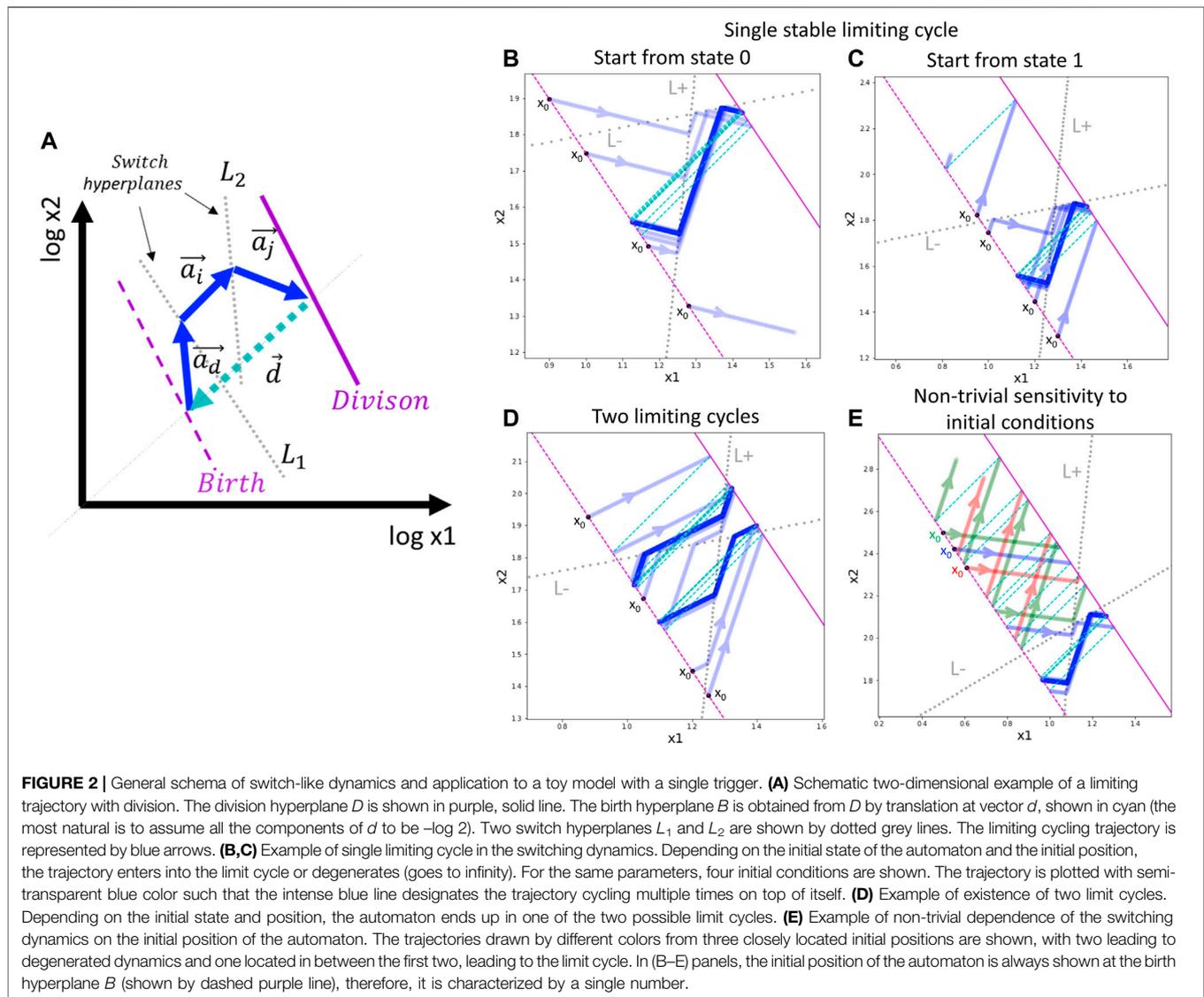
To summarize, the automaton A is characterized by its position and the intrinsic state, see **Figure 2A**. The asymptotic (in the infinite time limit) temporal dynamics of A is parameterized by a set of cell division planes $D = D_p$, $i = 1, \dots, k$, a set of switch functions $G = \{g_i, i = 1, \dots, p\}$, the corresponding switch hyperplanes $L = \{L_i, i = 1, \dots, p\}$, and the parameters of the cell division event (namely, the translation vector d and the state after cell division s_d).

It is convenient to encode the state s as a binary sequence of length r representing the on-off states of r triggers. In this case, a switch can be thought of as changing only one particular trigger from on to off or vice versa. In many situations, this makes the description of switch functions $g: S \rightarrow S$ quite natural as explained below. Also, the state of the trigger might not be strictly binary but characterized by several discrete positions, for example $\{0, 1, 2\}$, just as it is the case in modeling multi-level discrete dynamics, where each discrete variable can take a value from a pre-defined finite set of levels.

The exact asymptotic trajectory of the automaton A can, in principle, depend on the initial position x_0 and the initial intrinsic state s_0 of A .

3.3 Simple Example of Dynamics With Switches and Cell Division Events

In the above-described switch-like dynamics, one can find examples of relatively complex behaviors even for simple model settings (**Figure 2, B-E**). As an illustration, we modeled a simple dividing automaton characterized by a position vector x with only two coordinates x_1, x_2 . The automaton intrinsic state s encoded by only one binary trigger, so the automaton can be in two states $s = 0$ and $s = 1$, characterized by two vectors of movement a_0 and a_1 , respectively. In order to be able to modify the trigger in both directions, we have to introduce two switch hyperplanes $L^{(+)}$ and $L^{(-)}$ with corresponding switch functions $g^{(+)} = 1$ (switch trigger on) and $g^{(-)} = 0$ (switch trigger off). Note that in this case the switch functions are constant, i.e., they map any state (which can be either 0 or 1) to a particular state. Let us also assume that the division event changes the automaton position but does not change its intrinsic state.



In this simple toy example, by slightly varying parameters of the switching hyperplanes and the movement vectors, one can observe several interesting scenarios. Firstly, we observe that the automaton can approach and stay on a limit cycle trajectory, or it can diverge, meaning that one of the coordinates of the vector x goes to infinity or zero (Figure 2, B–C). Convergence or divergence to a limit cycle depends on the initial intrinsic state and the initial position of the automaton on the birth hyperplane.

In a more complex scenario, the switching dynamics trajectory can be characterized by two limit cycles that can be achieved from different initial intrinsic states and positions (Figure 2D).

By varying the positions of the switching hyperplanes in this toy example, one can observe the effect of non-trivial sensitivity to the initial conditions (Figure 2E). In this case, the birth hyperplane can be split into a sequence of alternating intervals of equal length such that starting from one interval, the dynamics

finally converges to the limit cycle, and starting from another interval, the dynamics diverges to infinity.

3.4 Two-Dimensional Model of Cell Cycle Progression, Fitted to the Single Cell Transcriptomic Data

Let us denote the aggregate signal related to the activation of genes associated with the S-phase of the cell cycle program as S , and the signal related to the activity of genes in G2 and M phases as M . Therefore, we will characterize the position of the automaton by a vector (x_S, x_M) , just as it is presented in Figure 1A, right panel. Let us denote the position of the turning points in the trajectory as $(x_S^{(i)}, x_M^{(i)})$, where $i \in \{0, A, B, C, D, E\}$.

We will encode the state of the system by the levels of two triggers, one associated with the S signal and another associated with the M signal. The three levels are denoted as a set

$\{2 = \text{synthesis}, 1 = \text{decay}, 0 = \text{degradation}\}$. Intuitively, these levels correspond to the state of active transcription of the corresponding set of transcripts (“synthesis”), absence of active transcription in which the transcripts are passively degraded according to some base rate (“decay”), and the process of active degradation when the transcripts are degraded more rapidly than the base rate (“degradation”). The state of the system is thus encoded by a pair of 3-level variables $i, j \in \{0, 1, 2\}$. The 2D vectors of linear movement a_{ij} are encoded by six rates $k_j^v, i \in \{0, 1, 2\}, v \in \{s, m\}$, such that $a_{ij} = (k_i^s, k_j^m)$. Following the intuition behind the introduced trigger levels, we assume constraints $k_2^v > 0, k_1^v < 0, k_0^v < 0, k_0^v < k_1^v < k_2^v$.

Let us introduce three switches. The first switch g_1 turns on the synthesis of both variables, i.e., $g_1(\bullet, \bullet) \rightarrow (2, 2)$, where \bullet designates any level of the trigger. The second switch turns off the synthesis of genes in S-phase: $g_2(2, \bullet) \rightarrow (1, \bullet)$. The third switch turns off all the transcription, $g_3(\bullet, \bullet) \rightarrow (1, 1)$. We assume that the division is possible only in the state (1, 1) with transcription switched off, and that after the division event, the cell enters into the state of active degradation of the cell cycle genes (0, 0).

The three introduced switches will be characterized by the corresponding switching hyperplanes. The first switch is triggered when the sum of the collective aggregated levels of expression of the genes involved in S and G2/M phases reaches some minimum c_{\min} , therefore, the linear functional associated with the first switch hyperplane is $f_1(x_s, x_m) = x_m + x_s - c_{\min}$. The second switch is triggered whenever the collective aggregated level of expression of S phase-associated genes reaches some maximum value S_{\max} , therefore, the linear functional associated with the second switch hyperplane is $f_2(x_s, x_M) = x_S - S_{\max}$. Finally, the third switch is triggered when the collective aggregated level of expression of G2/M phase-associated genes reaches some maximum value M_{\max} , therefore, the linear functional associated with the third switch hyperplane is $f_3(x_s, x_M) = x_M - M_{\max}$.

In the end, the cell division event is triggered when the collective aggregated level of expression of G2/M phase-associated genes crosses some threshold M_e , therefore, the linear functional associated with the division event is $f_d(x_s, x_M) = M_e - x_M$.

Let us define the number of parameters in this simple switching model. Three introduced switches are characterized by 4 parameters $c_{\min}, S_{\max}, M_{\max}, M_e$. There exist six rates k_i^v characterizing the movement vectors in the $9 = 3^2$ possible states, corresponding to all possible combinations of trigger levels. However, qualitatively, the dynamics in each automaton state is determined only by the direction of the corresponding vector and not its amplitude: therefore, one parameter per state visited is needed during the progression through the cell cycle. Under certain constraints on the rates formulated above, and also on the switch parameters (namely, $c_{\min} < S_{\max}, M_{\max}, M_e < M_{\max}$), the suggested model is constructed such that along the cell cycle trajectory only four states will be visited in a predefined order $(0, 0) \rightarrow (2, 2) \rightarrow (1, 2) \rightarrow (1, 1)$. Therefore, the total number of parameters equals 8.

Knowing the position of four characteristic points along the cell cycle trajectory, namely $(x_S^{(B)}, x_M^{(B)}), (x_S^{(C)}, x_M^{(C)}), (x_S^{(D)}, x_M^{(D)}), (x_S^{(E)}, x_M^{(E)})$, it is possible to completely parameterize the automaton. The starting and the ending point of the cell cycle trajectory must be connected by the relation $(x_S^{(0)}, x_M^{(0)}) = (x_S^{(E)}, x_M^{(E)}) + d$, where d is the vector with components $(-\log 2_{10}2, -\log 2_{10}2)$.

Therefore, we put $S_{\max} = x_S^{(C)}, M_{\max} = x_M^{(D)}, M_e = x_M^{(E)}$. Instead of using directly the B point, we will use the position of the non-proliferating cell with the maximum sum of the coordinates in the S, M plane, and we designate it as $x_S^{B'}, x_M^{B'}$ (other choices are also possible). Then $c_{\min} = x_S^{(B')} + x_M^{(B')}$. Then we define rates:

$$k_2^v = \frac{x_v^{(C)} - x_v^{(B')}}{\|x^C - x^{B'}\|}, k_1^S = \frac{x_S^{(C)} - x_S^{(D)}}{\|x^C - x^D\|}, k_1^M = \frac{x_M^{(E)} - x_M^{(D)}}{\|x^{(E)} - x^{(D)}\|}, k_0^v = \frac{x_v^{(B')} - x_v^{(0)}}{\|x^{(0)} - x^{B'}\|}$$

The resulting steady state cell cycle trajectory is shown in **Figure 3**.

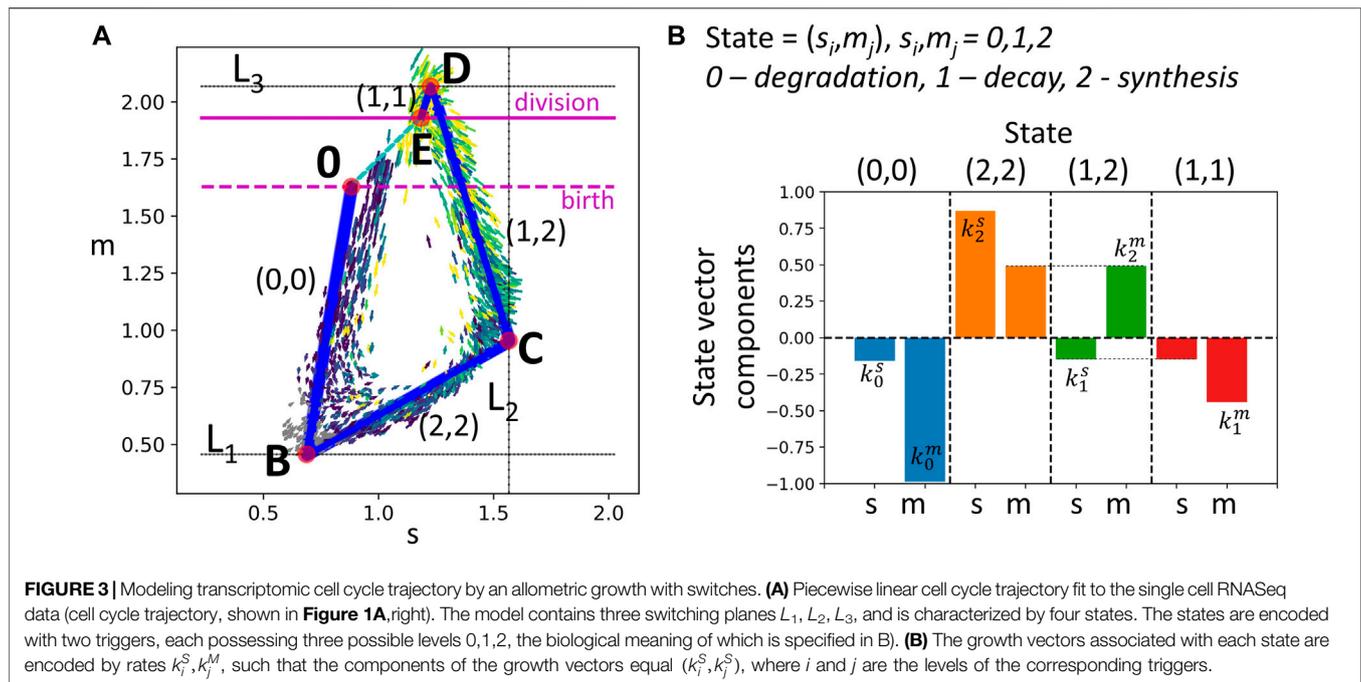
We denote the linear segments of the trajectory shown in **Figure 3** as T1, Ts, T2, Tm, assuming that they have significant overlap with G1, S, G2 and M phases correspondingly.

The suggested model describes 2D dynamics of the signals S, M which are empirically shown to explain most of the variance of all cell cycle genes in scRNASeq data (see below). However, higher-dimensional generalization of the suggested model is always possible. Also, in the model, we simplified the observed dynamics in **Figure 1A**, left which seems to contain five segments, with an additional curvature peak in point A. The segment A-B seems to contain non-proliferating cells, and might correspond to the transcriptional epoch most similar to the quiescent cell state, when the active degradation of the mitotic transcripts is completely finalized. The existence of this epoch is less pronounced in the S, M projection (**Figure 1A**, right), therefore we merged segments 0-A and A-B' as the first order approximation.

3.5 Connection Between the Effective Embedding Dimensionality of Cell Cycle Trajectory and the Number of Intrinsic States

The introduced cell cycle modeling framework is a simple and empirical model, lacking mechanistic details. Its main advantage is the possibility of analytical treatment of the most general geometrical cell cycle trajectory properties. In this section, we use this framework to prove a theorem connecting the number of the intrinsic states of the cell cycle trajectory and its intrinsic dimensionality.

This geometry is embedded into a space of omics measurements, whose dimensionality might be very high (e.g., expression of thousands of genes). However, we can assume that the intrinsic dimensionality (ID) of CCT is much smaller and that the extrinsic state of the cell progressing through the cell cycle can be characterized by n extensive variables, where n is relatively



small. We will refer to n as CCT embedding dimensionality. Empirically, it can be estimated by studying the snapshot of dividing single cells profiled with a particular technology, and computing its global intrinsic dimensionality (ID), provided that other non cell cycle-related sources of heterogeneity could be dismissed in measurements. Estimating ID can be done using one of the many existing methods for ID estimation (Albergante et al., 2019; Bac and Zinovyev, 2020; Bac et al., 2021).

Let us establish the expected relation between n and the number of intrinsic states m of the automaton approximating CCT. We intend to claim that theoretically n should match m under some natural assumptions.

We first state that m cannot be smaller than n . In the theory of allometric growth with switches this statement has a character of strict theorem (see below), $m \geq n$. Secondly, we state that n is expected to be at least equal to m . Both statements are based on argumentation using “general position” statements. However, the former one is strictly necessary, while the latter one represents a feasible hypothesis.

Theorem on the number of intrinsic cell cycle states. *The number of segments m in the cell cycle trajectory modeled by the automaton with switches and linear growth in logarithmic coordinates is not less than the cell cycle trajectory intrinsic dimensionality n , or $m \geq n$.*

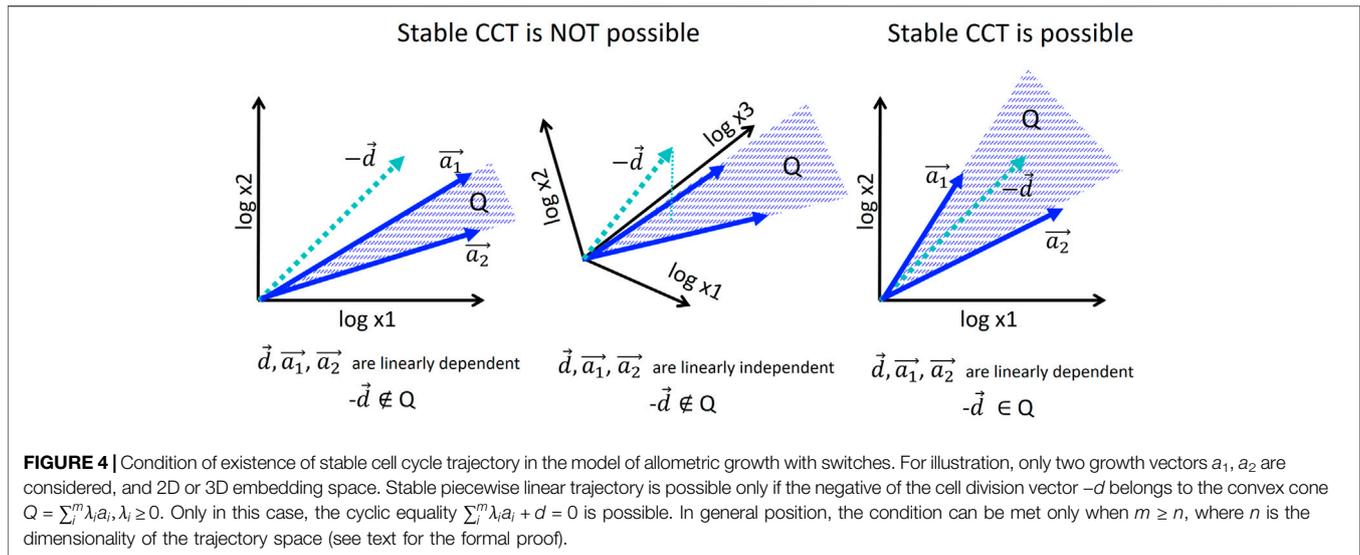
Proof. Let us consider the CCT dynamics in its n coordinates each of which represents an extensive variable. The variable extensiveness means, in particular, that its value, after the cell division moment, is divided by two. In logarithmic scale the cell division corresponds to the shift by vector $d \in \mathbb{R}^n$ with n coordinates each of which equals $-\log 2$. Each intrinsic state is associated with a growth vector $a_i \in \mathbb{R}^n$, $i = 1..m$. All non-negative linear combinations of a_i form a convex cone $Q = \{\sum_{i=1}^m \lambda_i a_i, \lambda_i \geq 0\}$. If $m < n$ then the set of vectors $\{d, \{a_i,$

$i = 1..m\}$ is almost always linear independent and $-d \notin Q$. Hence, $-d$ is linearly separable from Q , according to the standard separability theorems. Linear separability of a point from a convex cone can be expressed as that for any non-zero $x \in Q$ we can find a linear function $l(\cdot)$ such that $l(d) = 0$ and $l(x) > 0$. This makes the periodic cell cycle model impossible, because the function $l(x)$ increases along any growth direction, since for any i and $\lambda > 0$ we have $l(x + \lambda a_i) = l(x) + \lambda l(a_i) > l(x)$, and after cell division $l(\cdot)$ does not change since $l(x + d) = l(x) + l(d) = l(x)$. Therefore, the necessary condition of existence of stable cell cycle trajectory is $m \geq n$, when the set of vectors $\{d, \{a_i, i = 1..m\}\}$ is linearly dependent, and also such choice of a_i that $-d \in Q$. Only in this case one can satisfy the cyclic condition $\sum_i \lambda_i a_i + d = 0$ in general position of vectors $\{d, \{a_i, i = 1..m\}\}$. \square

In simple words, this means that if $m < n$ then in a general position, each cell division (shift by d) moves a cell state out of the subspace defined by the growth vectors. The only way to make the trajectory stay in this subspace is to make the cell division vector d belong to this subspace that can be guaranteed only if $m \geq n$ (see **Figure 4**). The condition $m \geq n$ is necessary but not sufficient for a model to converge to a limit cycle. For example, in **Figure 7**, $m = n = 2$ (the theorem condition is satisfied) but the limit cycle in the model can be achieved only from some initial conditions and for some choice of vectors a_0, a_1 .

Note that the proven Theorem is more general than the model of allometric growth with switches itself since it does not assume any particular shape of the switching surfaces L_k : they can be linear or nonlinear. Another generality consists in that the vector d can have any non-zero coordinates, not necessarily equal to $-\log 2$.

Examples in **Figures 2, 3** shows the case $n = 2, m > n$. The cell cycle trajectory modeled in **Figure 3** contains $m = 4$ segments in 2D, which makes the vectors $a_i \in \mathbb{R}^2$, $i = 1, \dots, 4$ linearly



dependent, and, of course, $d \in R^2$. The cell cycle model based on allometric growth is not contradictory in this case.

Now let us formulate our second statement. We can recall that vectors a_i are confined to the n -dimensional intrinsic subspace of CCT by projection from the multi-dimensional ambient space of all elementary measurements. The choice of n depends on our estimate of the CCT intrinsic dimensionality. However, movement along vectors a_i can be also seen in the complete space with thousands of coordinates. In this space, for sufficiently small m , any m vectors will almost always be linearly independent. Only projection into smaller than m -dimensional space will guarantee that these vectors are linearly dependent. This makes us hypothesize: if m segments are observed in CCT piecewise linear approximation in any linear projection then the most natural choice for n is at least m , i.e., $n \geq m$. Combining the two statements ($m \geq n$ and $n \geq m$) allows us to state that the correspondence $m = n$ is the most natural expectation for a cell cycle trajectory.

We explicitly verified this correspondence for the trajectory shown in **Figure 1**. The curvature analysis suggests the existence of five segments for the cell cycle trajectory reconstructed in the subspace of 30 first principal components of the complete dataset. However, some of these components might correspond to the variance not related to the progression through the cell cycle. In order to diminish the possible role of this variance, we considered a reduced version of the dataset confined to cell cycle-related genes only. We estimated the global intrinsic dimensionality, using six different linear ID estimators from scikit-dimension Python package (Bac et al., 2021), and it varied from 2 to 7, with average value 4.0. The scree plot shows existence of two dominant eigenvalues explaining 83% of total variance, indicating that the trajectory is relatively flat and located close to a 2D linear manifold. However, the residual variance demonstrated visible patterns related to transcriptional epochs in at least the first four principal components (**Figure 5**). The distribution of projections on the first four principal components well separated some transcriptional epochs (**Figure 5**, diagonal). Also, projections in

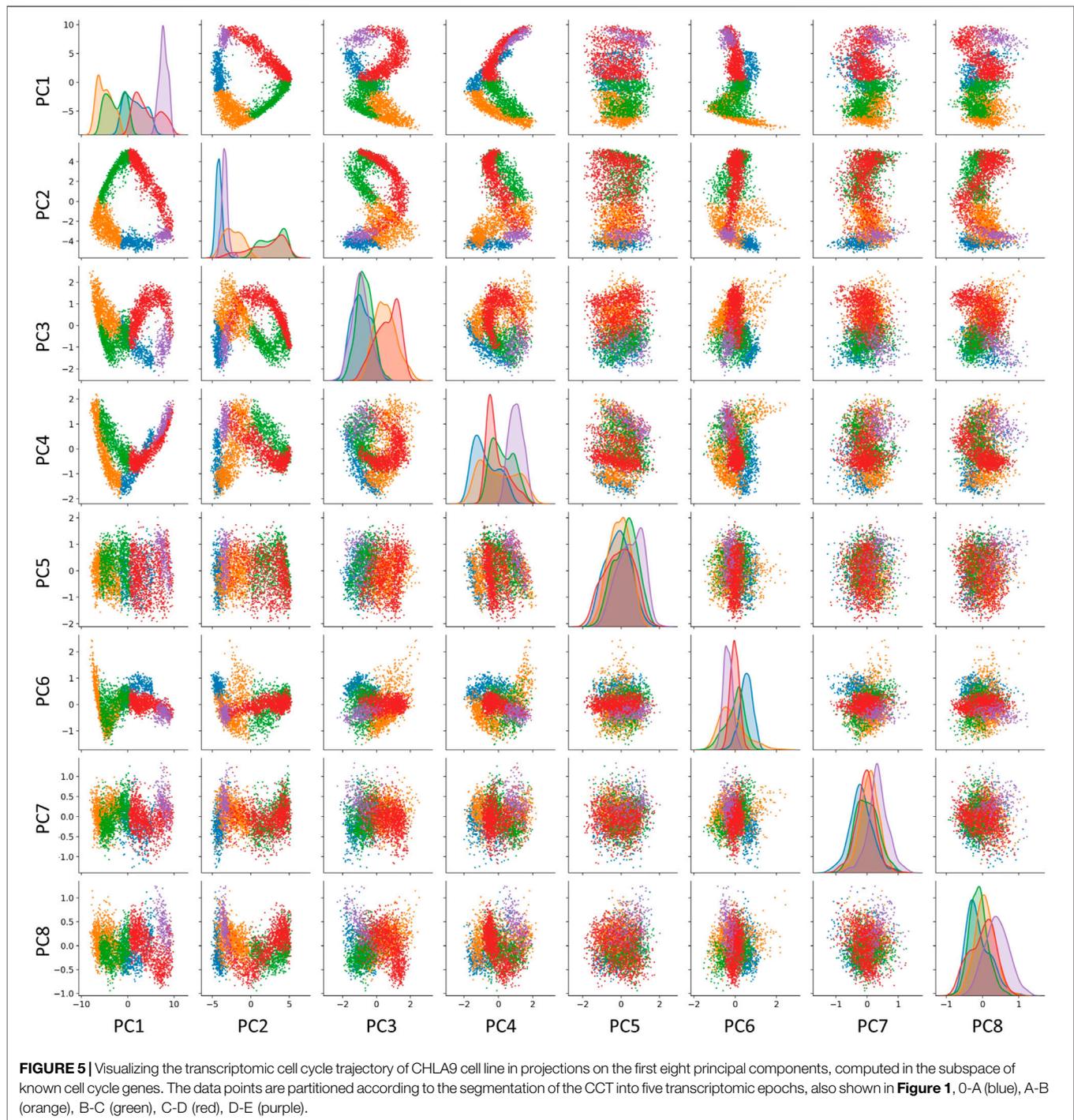
higher dimensions highlighted the existence of sharp turning points between the segments which were less clear in the 2D projection on the first two principal components.

In addition, we split the data points into five classes according to projection on five segments of the principal curve (0-A, A-B, B-C, C-D, D-E), each of which is approximately linear. For each of this class, we computed the unity vector corresponding to the direction of the first principal component in the space of cell cycle genes with 198 dimensions. Afterwards, we estimated the effective rank of the matrix composed of five vectors representing the directions of the transcriptional epochs in the multi-dimensional space (see Methods), and it appeared to be 4, which indicates to that at least four out of five vectors determining the trajectory segments can be considered linearly independent.

As a result, we concluded that the embedding dimensionality for the transcriptomic cell cycle trajectory can be estimated as close to four. Therefore, restricting the trajectory to the plane of aggregate collective expressions of genes associated with S phase and G2/M phase (which roughly corresponds to the first two principal components) is a useful but incomplete approximation of CCT dynamics. Our reasoning suggests searching for additional biologically meaningful and statistically independent scores describing the progression through the cell cycle. The concrete gene expression dynamics shown in **Figure 1B** provides a hint in this direction, but a careful and complete investigation of this question should be a subject of a separate study. As an additional argument, we can mention that some mathematical cell cycle models based on a fit to real data are four-dimensional (Singhania et al., 2011).

3.6 Extending the Modeling Formalism to Piecewise Smooth Trajectories: Simple Kinetic Model of Cell Cycle at Transcriptomic Level

The piecewise-linear model of automaton with switches described in the previous sections is phenomenological



and lacks any notion of physical time and connection to the underlying kinetics of the lumped expression of genes involved in S phase and G2/M phases. A simple way to make it more concrete but still analytically tractable consists in introducing explicit processes of synthesis and degradation of the corresponding quantities, with kinetic rates changing in time. The simplest form of such

dependence is piecewise-constant, with changes in the value of kinetic rates corresponding to the observed switches between transcriptional epochs of cell cycle progression.

Assuming the same epochs of cell cycle progression as above, and the same notations for variables (S , M , lumped expression of genes involved in S and G2/M phases correspondingly), their dynamics can be expressed as:

$$\begin{cases} \frac{dS}{dt} = k_t^S(t) - k_d^S(t)S \\ \frac{dM}{dt} = k_t^M(t) - k_d^M(t)M \end{cases} \quad (1)$$

These equations must be accompanied by circular boundary conditions

$$\begin{cases} S(T) = S_f S(0) \\ M(T) = M_f M(0) \end{cases} \quad (2)$$

where $S_f, M_f > 1$ are some numbers describing the drop of the lumped cell cycle variables after the moment of cell division. The most natural choice for them is $S_f, M_f = 2$, as before: however, here we prefer not to fix these parameters and rather fit them from the actually observed trajectory.

There exist several reasons for which S_f and M_f might appear in the range $1 \leq S_f, M_f \leq 2$ and not be equal. The most important of them is the technical biases introduced by sampling a limited amount of RNA, in the process of single cell transcriptome sequencing. It can lead to the situation when after cell division, the amount of RNA decreases non-uniformly between molecular processes. In particular, in all our experiments, we do observe the total amount of RNA reads does not decrease exactly by 2.0 and is rather close to 1.7-1.8. The decrease of the individual gene expression after cell division in terms of the number of reads, forms a bell-shaped distribution around this value with standard deviation close to 0.2.

Equation 1 with piecewise-constant in time kinetic rates and the boundary conditions 2) can be solved analytically for arbitrary number of levels in the piecewise-constant functions $k_t(t), k_d(t)$. The resulting dynamics in the plane $\log S(t), \log M(t)$ represents a cell cycle trajectory parameterized by physical time, which consists of piecewise-smooth segments of three types. If a segment is characterized by $k_t^S(t) = k_t^M(t) = 0$ then the corresponding segment is linear in the logarithmic coordinates (since the underlying dynamics is exponentially decaying). If a segment is characterized by $k_d^S(t) = k_d^M(t) = 0$ then the corresponding segment is also linear in both logarithmic and initial coordinates. For a segment where at least one degradation k_d^* and one production kinetic rate k_t^* are positive, the dynamics follows a nonlinear curve in the logarithmic space, which remains monotonous (each of the coordinates does not change the derivative sign). The nonlinearity of the segment becomes important when one of the variables is in a stage exponentially increasing or decreasing, while the other is in a linear or close to saturation stage. Otherwise, the segment remains close to a line in logarithmic coordinates.

In order to choose the number of constant levels of the kinetic rates, we studied the averaged RNA velocity values along the cell cycle as a function of pseudotime (see **Figures 6A,B**). For the S variable, we decided to keep only one non-zero level of $k_t^S(t)$ during the transcriptional epoch T_s , and two levels of $k_d^S(t)$, one for the exit from mitosis epoch and one for the rest of the dynamics. The choice was similar for M variable, but we took into account that a boost of expression of the

lumped $G2/M$ genes is visible in the beginning of the transcriptional epoch T_{2s} , just after switching off the S phase genes. During mitosis we assumed that all production rates are zero, corresponding to the lack of transcription in the M phase. The resulting choice of levels for the kinetic rates is shown in **Figure 6C**.

The advantage of the proposed simple model of cell cycle trajectory is that it is fully analytically tractable and its parameters can be uniquely fit to the cell cycle trajectory observed in single cell data, given some biologically meaningful constraints. Thus, assuming that the duration of mitosis is by order of magnitude faster than the T_{1s} epoch, for CHLA9 cell line one estimates the ratio between transcriptional epochs T_{2s} and T_{1s} close to 1.0 and the value of transcriptional boost of $G2/M$ genes in T_{2s} epoch close to 2.5-fold (**Figure 6C**). The determined values of all other parameters can be found in the Jupyter notebook at https://github.com/auranic/CellCycleTrajectory_SegmentModel.

3.7 Fitting Parameters of the Simple Kinetic Cell Cycle Model

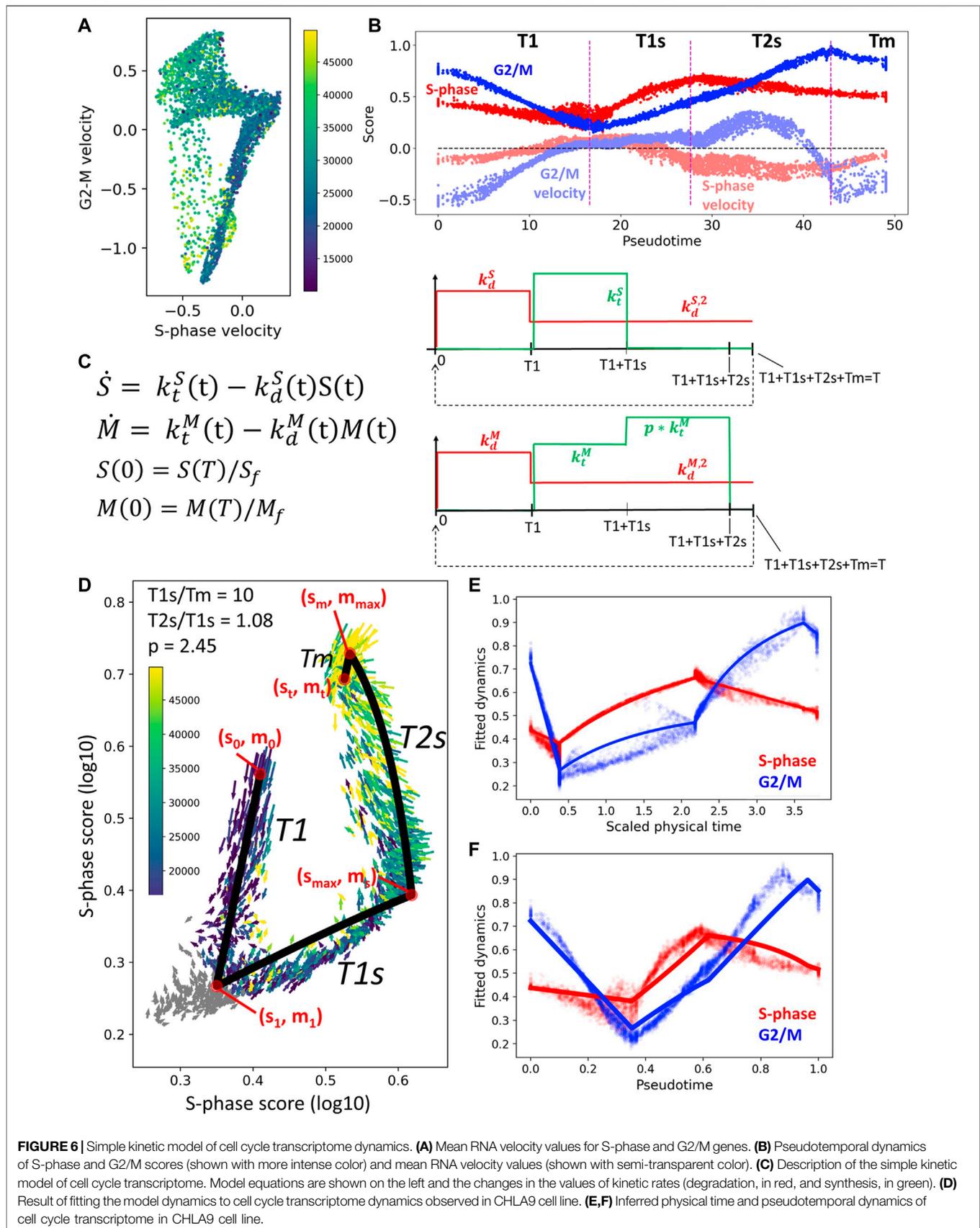
Using the choice of levels for piecewise constant kinetic rates shown in **Figure 6C**, we could derive the dependence of the initial state of the cell cycle from the kinetic rates and the durations of four transcriptional epochs T_1, T_{1s}, T_{2s}, T_m :

$$\begin{cases} S(0) = \frac{k_t^S}{k_d^{S,2}} \frac{e^{k_d^{S,2} T_{1s}} - 1}{S_f e^{k_d^{S,2} (T_{1s} + T_{2s} + T_m)} - e^{-k_d^S T_1}} \\ M(0) = \frac{k_t^M}{k_d^{M,2}} \frac{p \cdot e^{k_d^{M,2} (T_{1s} + T_{2s})} - (p-1)e^{k_d^{M,2} T_{1s}} - 1}{M_f e^{k_d^{M,2} (T_{1s} + T_{2s} + T_m)} - e^{-k_d^M T_1}} \end{cases} \quad (3)$$

Starting from the initial point of the trajectory $S(0), F(0)$ it is possible to analytically write down the coordinates of all other borders of the transcriptional epochs:

$$\begin{cases} S(T_1) = S(0)e^{-k_d^S T_1} \\ M(T_1) = M(0)e^{-k_d^M T_1} \\ S(T_1 + T_{1s}) = \frac{k_t^S}{k_d^{S,2}} \left(1 - \left(1 - \frac{k_d^{S,2}}{k_t^S} S(T_1) \right) e^{-k_d^{S,2} T_{1s}} \right) \\ M(T_1 + T_{1s}) = \frac{k_t^M}{k_d^{M,2}} \left(1 - \left(1 - \frac{k_d^{M,2}}{k_t^M} M(T_1) \right) e^{-k_d^{M,2} T_{1s}} \right) \\ S(T_1 + T_{1s} + T_{2s}) = S(T_1 + T_{1s}) e^{-k_d^{S,2} T_{2s}} \\ M(T_1 + T_{1s} + T_{2s}) = \frac{p \cdot k_t^M}{k_d^{M,2}} \left(1 - \left(1 - \frac{k_d^{M,2}}{p \cdot k_t^M} M(T_1 + T_{1s}) \right) e^{-k_d^{M,2} T_{2s}} \right) \\ S(T) = S(T_1 + T_{1s} + T_{2s}) e^{-k_d^S T_m} \\ M(T) = M(T_1 + T_{1s} + T_{2s}) e^{-k_d^M T_m} \end{cases} \quad (4)$$

where $T = T_1 + T_{1s} + T_{2s} + T_m$ is the full duration of the cell cycle. One can estimate the position of these points from the analysis of observed cell cycle trajectory curvature ($(s_0, m_0), (s_1, m_1), (s_{\max}, m_s), (s_m, m_{\max}), (s_t, m_t)$), shown by red points in **Figure 6D**) by requiring that the model trajectory should pass as close as possible to them. This defines an optimization problem which can be easily solved numerically by iterations, using the simplest



fixed-point algorithm. The details of parameter fitting are provided in the Jupyter notebook at https://github.com/auranic/CellCycleTrajectory_SegmentModel.

We note that this optimization does not allow us to determine all the model parameters uniquely, since they enter in the aforementioned optimization functional as certain combinations (as simple rational functions), namely, $\frac{k_i^S}{k_d^{S,2}}, \frac{k_i^M}{k_d^{M,2}}, k_d^S T_1, k_d^M T_1, k_d^{S,2} T_{1s}, k_d^{M,2} T_{1s}, k_d^{S,2} T_{2s}, k_d^{M,2} T_{2s}, k_d^{S,2} T_m, k_d^{M,2} T_m$. Two other parameters M, S define the observed cell division vector in (2). One extra parameter p denotes transcriptional production acceleration of G2/M genes during the transcriptional epoch T2s compared to the transcriptional epoch T1s (Figure 6C). Not all these quantities are independent, some of them are connected through nonlinear relations:

$$\frac{k_d^{S,2} \cdot T_{2s}}{k_d^{S,2} \cdot T_{1s}} = \frac{k_d^{M,2} \cdot T_{2s}}{k_d^{M,2} \cdot T_{1s}}, \frac{k_d^{S,2} \cdot T_m}{k_d^{S,2} \cdot T_{1s}} = \frac{k_d^{M,2} \cdot T_m}{k_d^{M,2} \cdot T_{1s}}, \quad (5)$$

which overall gives 11 independent combinations of parameters provided 10 measurable coordinates of cell trajectory turning points in Figure 6D.

Altogether, this means that 1) one needs to introduce at least one additional constraint in order to make the trajectory reconstruction unique and 2) physical time of the epochs T_1, T_{1s}, T_{2s}, T_m can not be uniquely computed from the cell cycle trajectory observed in the plane of S-, G2/M-phase scores. From the analysis of Eq. 5 it follows that the model can be uniquely parameterized if one will constrain one of the three quantities $p, \frac{T_{2s}}{T_{1s}}, \frac{T_m}{T_{1s}}$. Finally, it is convenient to fix the durations T_1, T_{1s} to some arbitrary values which allows to determine parameters k_d^S, k_d^M and the ratios $\frac{T_{2s}}{T_{1s}}, \frac{T_m}{T_{1s}}$.

In our numerical experiments, we fixed the values of T_1 and T_{1s} to their corresponding pseudotemporal durations (as the corresponding fractions of the total length of the cell cycle trajectory). We also fixed the ratio $\frac{T_m}{T_{1s}} = 10$, assuming that the mitosis must be fast in physical time compared to the transcriptional epoch including activating the expression of the genes involved in the S-phase.

3.8 Simulating Cell Cycle Trajectories With Various Durations of Temporal Transcriptional Epochs

After fitting the kinetic parameters for an observable in the S-phase vs G2/M score plane cell cycle trajectory, one can perturb the parameters and investigate how the trajectory geometry depends on them.

In real life scRNASeq datasets, we observe that CCT geometry can appear very different in various biological systems. When projecting onto the plane of standard scores of S-phase and G2/M phase genes, scRNASeq datasets might not always reveal the circular nature of CCT. In some cases, the circular structure is not at all detectable via this projection, (Figure 7), and the two scores might be connected via a strong positive or negative correlation. Also, in some systems we observed co-existence of several CCT shapes, like it is the case in the U2OS cell line dataset

(GSE146773). The univariate histograms of two score distributions might be characterized by bi- or uni-modal character.

Quite strikingly, we were able to reproduce these patterns qualitatively by fitting the kinetic parameters to the CHLA9 scRNASeq dataset, and then by manipulating the durations of T1, T1s and T2s transcriptional epochs and producing computer-simulated trajectory examples. Thus, significant reduction in the duration of both T1 and T1s epochs led to the negative correlation pattern between S-phase and G2/M scores. This could be interpreted as drastic reduction of the G1 cell cycle phase. In real life datasets, such pattern has been observed in human embryonic stem cells (dataset GSE85917).

If both T1 and T2s were shortened then this led to the increase of the positive correlation between two scores, (Figure 7). This pattern was indeed observed in human bone marrow and human neural epithelial stem cell-related single cell datasets (GSE99095 and GSE81475).

3.9 Predicting Cell Line Doubling Time From the Geometrical Properties of Cell Cycle Trajectory

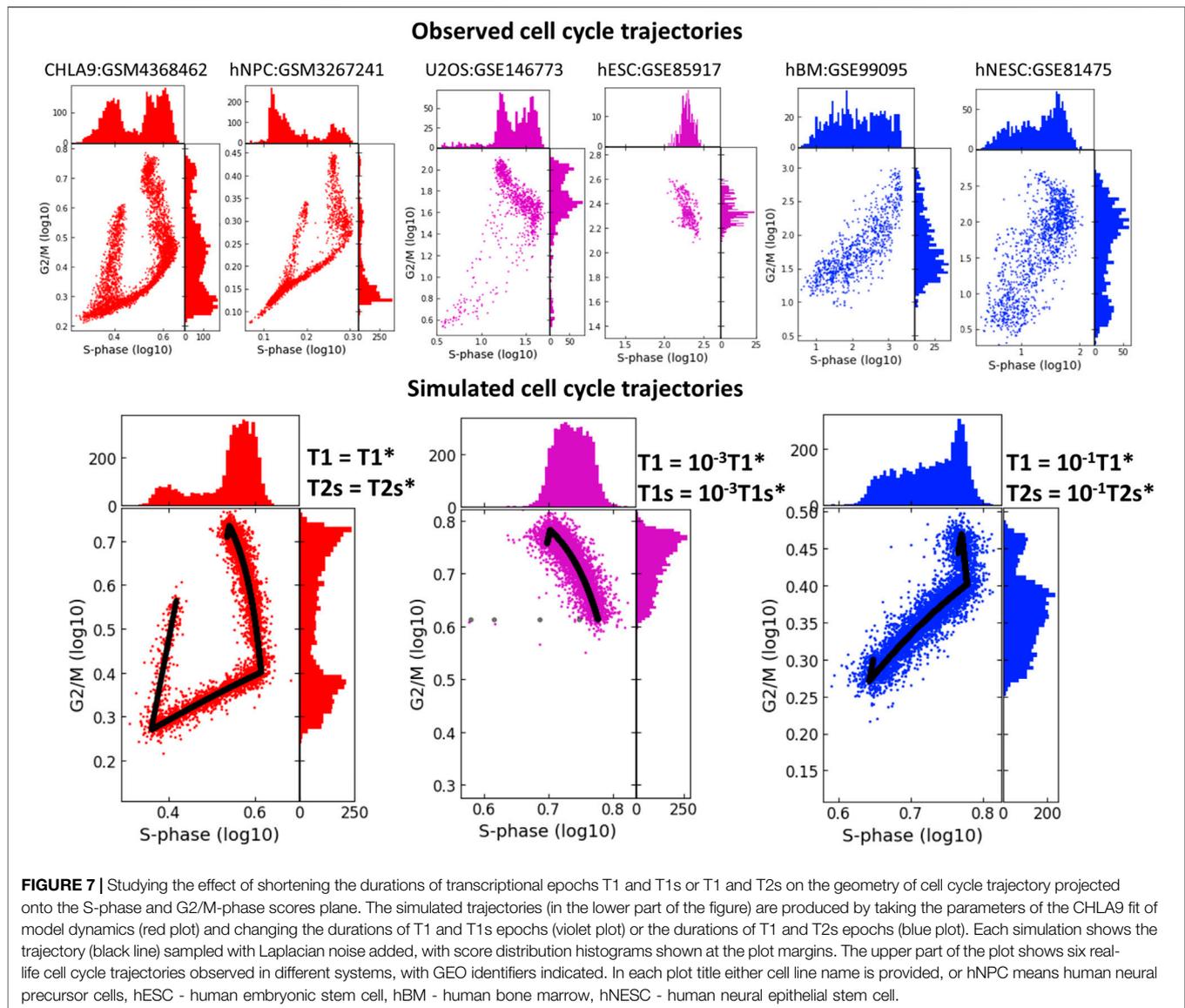
The developed simple kinetic model leads to a simple prediction which can be validated: the total length of the transcriptomic cell cycle trajectory must diminish in rapidly dividing cells. This can be interpreted as a consequence of the fact that in a rapid proliferation process, during the post-mitotic G1 phase (T1 transcriptional epoch), there is not enough time to degrade all mitotic transcripts produced before the cell division moment, so they are reused in the consequent cell cycle phases, shortening the subsequent G1 phase.

We verified this prediction in a relatively large collection of cell line scRNASeq datasets. Using the data from Cellosaurus database, we identified those few ones for which the cell line doubling time has been estimated, and for which the number of available good quality single cell profiles exceeded 300.

We used the total length of the principal circle fit in the 2D plane of the scaled to maximum equals 1 cell cycle phase scores, as a proxy to quantify the level of CCT contraction (see Methods). This measure was correlated with cell line doubling time in hours. Two cell lines CHLA10 and SCC25 appeared to be strong outliers from otherwise significant positive regression line (Pearson correlation 0.931, p -value = 10^{-5}) (Figure 8). When this regression line was used as a predictor, CHLA10 cell line was predicted to have doubling time around 64 h (instead of determined by database search of around 32 h) and for SCC25 around 78 instead of 50 h. It is known that cell line doubling time can vary depending on the growth conditions, so we hypothesized that this variability could explain the appearance of two outliers. If two of them were kept in the regression calculation, it remained significant but less strong (Pearson correlation 0.67, p -value = 0.01).

3.10 Code Availability

The Python notebooks allowing the reader to reproduce all the computations presented in this manuscript are freely available



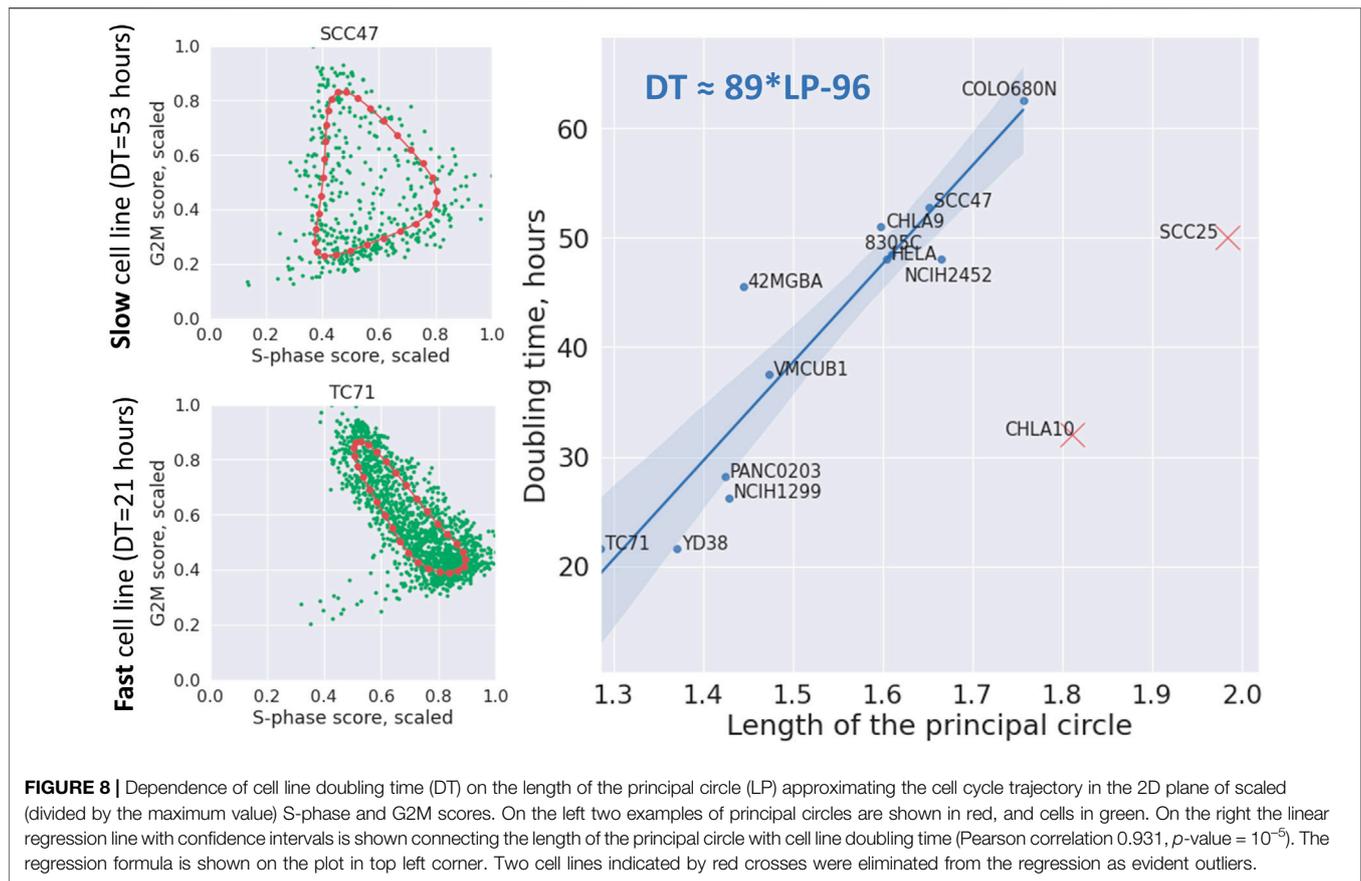
from https://github.com/auranic/CellCycleTrajectory_SegmentModel.

4 DISCUSSION

This paper provides a framework for analyzing the cell cycle trajectories using single cell omics measurements such as scRNASeq data. Unlike the previously suggested model of the trajectory as a flat circle, we provide arguments that at least in some conditions the piecewise-linear in logarithmic coordinates approximation appears to fit the single cell transcriptomic data and to be biologically tractable. In particular, it allows us to delineate transcriptional epochs of cell cycle at which the corresponding segment of the trajectory remains close to linear in logarithmic coordinates which corresponds to locally allometric changes of the transcriptome.

We suggest two modeling formalisms to recapitulate the cell cycle transcriptomic dynamics as a sequence of switches. The first one is purely phenomenological and describes the dynamics as a change of states of a hidden automaton, leading to the switches of parameters of allometric growth, followed by a shift representing the cell division event. The advantage of this formalism is that it allows us to treat most general properties of cell cycle trajectory geometry.

In particular, we could prove a fundamental theorem on the number of intrinsic cell cycle states, which connects the number of linear segments in the trajectory and the embedding dimensionality of the cell cycle trajectory. The nature of this theorem, relying on “general position”-type arguments, is reminiscent of the well-known results imposing constraints on the number of the system’s internal states and the effective dimensionality of its environment, in several fields of science. For example, the Gause’s law of competitive exclusion and its



generalizations states that the number of competing species is limited by the effective number of resources, characterizing the environment (Gauze, 1934; Gorban, 2007). The famous Gibbs' phase rule in thermodynamics connects the effective number of the intensive variables with the number of components and phases in a system at thermodynamic equilibrium (Gibbs, 1961; Alper, 1999). All these results are also similar in terms of practical difficulties related to determining the effective system's dimensionality.

From the physico-chemical point of view, the effective dimensionality is the number of the substances "lumps" in the cell cycle kinetics. Lumping-analysis produces a partition of all chemical species into a few groups and then considers these groups ("lumps") as independent entities (Wei and Kuo, 1969). "Amounts" of these lumps are the combinations of the amounts of the chemical species (Li and Rabitz, 1989; Li and Rabitz, 1990). The theorem on the number of intrinsic cell cycle states that the number of lumps n does not exceed the number of the internal states of the cell cycle transcription machinery. This means that kinetics allows reduction of the huge-dimensional space of all components to $n \leq m$ number of aggregated lumps.

The second modeling formalism that we suggested connects the geometric properties of the cell cycle trajectory to the underlying transcriptional kinetics and physical time. It uses the simplest chemical kinetics equations with kinetic rates represented as piecewise-constant functions of time. We show

that the suggested model is fully analytically tractable and, under some biologically transparent assumptions, allows unique determination of its independent parameter combinations. This type of modeling allowed us to explicitly study the relation between pseudotime and physical time.

The precise connection between physical time and pseudotime (geometric time) in the cell cycle is worth studying in more detail since this is the central question in the dynamic phenotyping approach in general (Golovenkin et al., 2020). Some of these relations can be potentially quantified from exploring the variations of point density along the inferred trajectories (Chen et al., 2019). Related to this, one can expect non-trivial phenomena in studying the cell cycle trajectory, such as effects of partial cell population synchronization under assumption of equal cell cycle durations in individual cells. This effect can lead to the appearance of density peaks in the reconstructed cell cycle trajectories that cannot be explained by nonlinear relation between physical time and pseudotime (Gorban, 2007).

As one of the applications of the suggested modeling formalism, we performed several numerical experiments on changing the durations of the transcriptional epochs overlapping with G1 or G2 cell cycle phases. We observed that these parameters might have a drastic effect on the shape of the CCT geometry and the form of the univariate variable distributions. This model prediction can be qualitatively confirmed by observing CCT properties of several *in vitro* and *in vivo* systems. The effect of CCT shrinkage might be

relevant in characterizing the cell cycle properties in various conditions: for example, when one can manipulate the activity of an oncogene (Aynaud et al., 2020). We show that the CCT geometry can be predictive to estimate the cell line doubling time which can be a proxy of cell cycle duration.

The relation between transcriptomic dynamics and the established definitions of cell cycle phases and cell cycle checkpoints has been discussed and even quantified using standard molecular biology techniques (Giotti et al., 2019; Hsiao et al., 2020). In this study, we deliberately leave open the question on defining the exact cell cycle phase borders from the transcriptomic CCT geometry. We found that this relation can not be the exact match: one of the reasons for this is delayed production of proteins, and dependence of the cell cycle progression from post-translational protein modifications. The transcriptomic dynamics is relatively slow, and activation of protein synthesis is switched on in advance, leaving time for producing enough proteins needed at a certain stage of the cell cycle molecular program. Same is true for the process of degradation of RNAs involved in cell cycle: a cell needs enough time after mitosis to degrade all cell cycle-related transcripts.

The suggested formalism is not limited to transcriptomic data. It looks promising to analyze the geometrical properties of cell cycle trajectory measured in unsynchronized cell populations profiled at various levels of molecular description, including epigenetics and protein expression, when the datasets of sufficient volume and quality will become available.

A more mechanistic description of the cell cycle has been already proposed in the context of yeast or mammalian cells (Tyson, 1991; Novák and Tyson, 2004). The mathematical models can be based on chemical kinetics or on discrete or hybrid frameworks but in all cases, the difficulty when constructing these models is to select the genes that can capture the main features of the cell cycle and the different

events that allow the switch from one phase to another. We anticipate that the type of analyses presented here could orient the choice of these genes and inform on their dynamics.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article. The processed scRNA-Seq data used in this study are available from <http://doi.org/10.5281/zenodo.5017357>. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AZ, MS, and AG designed the study and performed analytical calculations. AZ, AF, CG, and AC contributed to computer simulations and code development. AZ, AC, AF, and CG participated in data collection and analysis. LC, EB, AC, and CG interpreted the analysis results. AZ and MS wrote the manuscript draft. All authors read the manuscript and edited it.

FUNDING

This work was supported by the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute), The work was supported by the Ministry of Science and Higher Education of the Russian Federation (Project No. 075-15-2021-634) and by European Union’s Horizon 2020 program (grant No. 826121, iPC project). These funding sources had no role in the design, execution and interpretation of the results of this study.

REFERENCES

- Albergante, L., Bac, J., and Zinovyev, A. (2019). “Estimating the Effective Dimension of Large Biological Datasets Using Fisher Separability Analysis,” in *Proceedings of the International Joint Conference on Neural Networks*. doi:10.1109/ijcnn.2019.8852450
- Albergante, L., Mirkes, E., Bac, J., Chen, H., Martin, A., Faure, L., et al. (2020). Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via ELPiGraph. *Entropy* 22 (3), 296. doi:10.3390/e22030296
- Alper, J. S. (1999). The Gibbs Phase Rule Revisited: Interrelationships between Components and Phases. *J. Chem. Educ.* 76 (11), 1567–1569. doi:10.1021/ed076p1567
- Altinok, A., Lévi, F., and Goldbeter, A. (2007). A Cell Cycle Automaton Model for Probing Circadian Patterns of Anticancer Drug Delivery. *Adv. Drug Deliv. Rev.* 59 (9–10), 1036–1053. doi:10.1016/J.ADDR.2006.09.022
- Aynaud, M.-M., Mirabeau, O., Gruel, N., Grossetête, S., Boeva, V., Durand, S., et al. (2020). Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cel Rep.* 30 (6), 1767–1779. doi:10.1016/j.celrep.2020.01.049
- Azizi, E., Carr, A. J., Plitas, G., Cornish, A. E., Konopacki, C., Prabhakaran, S., et al. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* 174 (5), 1293–1308. doi:10.1016/j.cell.2018.05.060
- Bac, J., Mirkes, E. M., Gorban, A. N., Tyukin, I., and Zinovyev, A. (2021). Scikit-dimension: a Python Package for Intrinsic Dimension Estimation. *Entropy Basel* 23:1368. doi:10.3390/e23101368
- Bac, J., and Zinovyev, A. (2020). Lizard Brain: Tackling Locally Low-Dimensional yet Globally Complex Organization of Multi-Dimensional Datasets. *Front. Neurobot.* 13. doi:10.3389/fnbot.2019.00110
- Bairoch, A. (2018). The Cellosaurus, a Cell-Line Knowledge Resource. *J. Biomol. Tech.* 29 (2), 25–38. doi:10.1038/s41588-020-00726-610.7171/jbt.18-2902-002
- Bauer, E. (1935). *Theoretical Biology*. Budapest: Akademiai Kiado.
- Bernard, D., Mondesert, O., Gomes, A., Duthen, Y., Lobjois, V., Cussat-Blanc, S., et al. (2019). A Checkpoint-Oriented Cell Cycle Simulation Model. *Cell CycleZIP* 18 (8), 795–808. doi:10.1080/15384101.2019.1591125/SUPPL_FILE/KCCY_A_1591125_SMI721
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal Probabilistic RNA-Seq Quantification. *Nat. Biotechnol.* 34 (5), 525–527. doi:10.1038/nbt.3519
- Chen, H., Albergante, L., Hsu, J. Y., Lareau, C. A., Lo Bosco, G., Guan, J., et al. (2019). Single-cell Trajectories Reconstruction, Exploration and Mapping of Omics Data with STREAM. *Nat. Commun.* 10 (1). doi:10.1038/s41467-019-09670-4
- Chen, K. C., Calzone, L., Csikasz-Nagy, A., Cross, F. R., Novak, B., and Tyson, J. J. (2004). Integrative Analysis of Cell Cycle Control in Budding Yeast. *MBoC* 15 (8), 3841–3862. doi:10.1091/mbc.E03-11-0794
- Chen, K. C., Calzone, L., Csikasz-Nagy, A., Cross, F. R., Novak, B., and Tyson, J. J. (2004). Integrative Analysis of Cell Cycle Control in Budding Yeast. *MBoC* 15 (8), 3841–3862. doi:10.1091/mbc.e03-11-0794
- Deritei, D., Rozum, J., Ravasz Regan, E., and Albert, R. (2019)201. A Feedback Loop of Conditionally Stable Circuits Drives the Cell Cycle from Checkpoint to Checkpoint. *Sci. Rep.* 9 (11), 1–19. doi:10.1038/s41598-019-52725-1

- Dominguez, D., Tsai, Y.-H., Gomez, N., Jha, D. K., Davis, I., and Wang, Z. (2016). A High-Resolution Transcriptome Map of Cell Cycle Reveals Novel Connections between Periodic Genes and Cancer. *Cell Res* 26 (8), 946–962. doi:10.1038/cr.2016.84
- Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical Analysis of a Generic Boolean Model for the Control of the Mammalian Cell Cycle. *BioinformaticsOxford Acad.* 22 (14), e124–31. doi:10.1093/bioinformatics/btl210
- Gauze, G. F. (1934). *The Struggle for Existence*. Baltimore: The Williams I& Wilkins company. Available from: <https://www.biodiversitylibrary.org/item/23409>.
- Gibbs, J. W. (1961). *The Scientific Papers*, 1. New York: Thermodynamics. Dover.
- Giotti, B., Chen, S.-H., Barnett, M. W., Regan, T., Ly, T., Wiemann, S., et al. (2019). Assembly of a Parts List of the Human Mitotic Cell Cycle Machinery. *J. Mol. Cell Biol.* 11 (8), 703–718. doi:10.1093/jmcb/mjy063
- Golovenkin, S. E., Bac, J., Chervov, A., Mirkes, E. M., Orlova, Y. V., Barillot, E., et al. (2020). Trajectories, Bifurcations, and Pseudo-time in Large Clinical Datasets: Applications to Myocardial Infarction and Diabetes Data. *GigaScience* 9 (11), 1–20. doi:10.1093/gigascience/giaa128
- Gorban, A. N. (2018). Model Reduction in Chemical Dynamics: Slow Invariant Manifolds, Singular Perturbations, Thermodynamic Estimates, and Analysis of Reaction Graph .
- Gorban, A. N. (2007). Selection Theorem for Systems with Inheritance. *Math. Model. Nat. Phenom.* 2 (4), 1–45. doi:10.1051/mmnp:2008024
- Holford, N. H. G., and Anderson, B. J. (2017). Allometric Size: the Scientific Theory and Extension to normal Fat Mass. *Eur. J. Pharm. Sci.* 109, S59–S64. doi:10.1016/j.ejps.2017.05.056
- Hsiao, C. J., Tung, P., Blischak, J. D., Burnett, J. E., Barr, K. A., Dey, K. K., et al. (2020). Characterizing and Inferring Quantitative Cell Cycle Phase in Single-Cell RNA-Seq Data Analysis. *Genome Res.* 30 (4), 611–621. doi:10.1101/gr.247759.118
- Hunt, T. (1991). Cell Cycle Gets More Cyclins. *Nature* 350 (6318), 462–463. doi:10.1038/350462a0
- Hunt, T., Nasmyth, K., and Novák, B. (2011). The Cell Cycle. *Phil. Trans. R. Soc. B* 366 (1584), 3494–3497. doi:10.1098/rstb.2011.0274
- Ingolia, N. T., and Murray, A. W. (2004). The Ups and downs of Modeling the Cell Cycle. *Curr Biol*, 14 R771–R777. doi:10.1016/j.cub.2004.09.018
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 48 (D1), D498–D503. doi:10.1093/nar/gkz1031
- Kinker, G. S., Greenwald, A. C., Tal, R., Orlova, Z., Cuoco, M. S., McFarland, J. M., et al. (2020). Pan-cancer Single-Cell RNA-Seq Identifies Recurring Programs of Cellular Heterogeneity. *Nat. Genet.* 52 (11), 1208–1218. doi:10.1038/s41588-020-00726-6
- Kotani, S. (2002). A Computational Model of Mammalian Cell Cycle Using Petri Nets. *Genome Inform.* 460, 459–460. doi:10.11234/GI1990.13.459
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., et al. (2018). RNA Velocity of Single Cells. *Nature* 560 (7719), 494–498. doi:10.1038/s41586-018-0414-6
- Leng, N., Chu, L.-F., Barry, C., Li, Y., Choi, J., Li, X., et al. (2015). Oscope Identifies Oscillatory Genes in Unsynchronized Single-Cell RNA-Seq Experiments. *Nat. Methods* 12 (10), 947–950. doi:10.1038/nmeth.3549
- Li, G., and Rabitz, H. (1990). A General Analysis of Approximate Lumping in Chemical Kinetics. *Chem. Eng. Sci.* 45 (4), 977–1002. doi:10.1016/0009-2509(90)85020-E
- Li, G., and Rabitz, H. (1989). A General Analysis of Exact Lumping in Chemical Kinetics. *Chem. Eng. Sci.* 44 (6), 1413–1430. doi:10.1016/0009-2509(89)85014-6
- Liang, S., Wang, F., Han, J., and Chen, K. (2020). Latent Periodic Process Inference from Single-Cell RNA-Seq Data. *Nat. Commun.* 11 (1441), 11. doi:10.1038/s41467-020-15295-9
- Liu, Z., Lou, H., Xie, K., Wang, H., Chen, N., Aparicio, O. M., et al. (2017). Reconstructing Cell Cycle Pseudo Time-Series via Single-Cell Transcriptome Data. *Nat. Commun.* 8 (1), 1–9. doi:10.1038/s41467-017-00039-z
- Mahdessian, D., Cesnik, A. J., Gnann, C., Danielsson, F., Stenström, L., Arif, M., et al. (2021). Spatiotemporal Dissection of the Cell Cycle with Single-Cell Proteogenomics. *Nature* 590 (7847), 649–654. doi:10.1038/s41586-021-03232-9
- Miller, H. E., Gorthi, A., Bassani, N., Lawrence, L. A., Iskra, B. S., and Bishop, A. J. R. (2020). Reconstruction of ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of Ewsr1-Fli1 Permissibility. *Cancers* 12 (4), 948. doi:10.3390/cancers12040948
- Noel, V., Grigoriev, D., Vakulenko, S., and Radulescu, O. (2012). Tropical Geometries and Dynamics of Biochemical Networks Application to Hybrid Cell Cycle Models. *Electron. Notes Theor. Comp. Sci.* 284, 75–91. doi:10.1016/j.entcs.2012.05.016
- Noël, V., Vakulenko, S., and Radulescu, O. (2013). A Hybrid Mammalian Cell Cycle Model. *Electron. Proc. Theor. Comput. Sci.* 125, 68–83. doi:10.4204/EPTCS.125.5
- Novák, B., and Tyson, J. J. (2004). A Model for Restriction point Control of the Mammalian Cell Cycle. *J. Theor. Biol.* 230 (4), 563–579. doi:10.1016/j.jtbi.2004.04.039
- Packard, G. C. (2017). The Essential Role for Graphs in Allometric Analysis. *Biol. J. Linn. Soc.* 120 (2), 468–473.
- Pretzsch, H. (2020). The Course of Tree Growth. Theory and Reality. *For. Ecol. Manag.* 478, 118508. doi:10.1016/j.foreco.2020.118508
- Radulescu, O., Gorban, A. N., Zinovyev, A., and Noel, V. (2012). Reduction of Dynamical Biochemical Reactions Networks in Computational Biology. *Front. Gene* 3, 131. doi:10.3389/fgene.2012.00131
- Schwabe, D., Formichetti, S., Junker, J. P., Falcke, M., and Rajewsky, N. (2020). The Transcriptome Dynamics of Single Cells during the Cell Cycle. *Mol. Syst. Biol.* 16 (11), e9946. doi:10.15252/msb.20209946
- Shkolnik, E. M. (1989). “Dynamic Theory of Cell Cycle,” in *Dynamics of Chemical and Biological Systems* (Russian: Nauka plc. Siberian branch), 159–190.
- Sible, J. C., and Tyson, J. J. (2007). Mathematical Modeling as a Tool for Investigating Cell Cycle Control Networks. *Methods* 41 (2), 238–247. doi:10.1016/j.ymeth.2006.08.003
- Singhania, R., Sramkoski, R. M., Jacobberger, J. W., and Tyson, J. J. (2011). A Hybrid Model of Mammalian Cell Cycle Regulation. *Plos Comput. Biol.* 7 (2), e1001077. doi:10.1371/JOURNAL.PCBI.1001077
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., et al. (2016). Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq. *Science* 352 (6282), 189–196. doi:10.1126/science.aad0501
- Tyson, J. J. (1991). Modeling the Cell Division Cycle: Cdc2 and Cyclin Interactions. *Proc. Natl. Acad. Sci.* 88 (16), 7328–7332. doi:10.1073/pnas.88.16.7328
- Tzamali, E., Tzedakis, G., and Sakkalis, V. (2020). Modeling How Heterogeneity in Cell Cycle Length Affects Cancer Cell Growth Dynamics in Response to Treatment. *Front. Oncol.* 10, 1552. doi:10.3389/FONC.2020.01552/BIBTEX
- Wei, J., and Kuo, J. C. W. (1969). Lumping Analysis in Monomolecular Reaction Systems. Analysis of the Exactly Lumpable System. *Ind. Eng. Chem. Fund.* 8 (1), 114–123. doi:10.1021/i160029a019
- White, C. R., Marshall, D. J., Alton, L. A., Arnold, P. A., Beaman, J. E., Bywater, C. L., et al. (2019). The Origin and Maintenance of Metabolic Allometry in Animals. *Nat. Ecol. Evol.* 3 (4), 598–603. doi:10.1038/s41559-019-0839-9
- Zetterberg, A., Larsson, O., and Wiman, K. G. (1995). What Is the Restriction point? *Curr. Opin. Cell Biol* 7, 835–842. doi:10.1016/0955-0674(95)80067-0
- Zhou, X., Yang, M., Liu, Z., Li, P., Xie, B., and Peng, C. (2021). Dynamic Allometric Scaling of Tree Biomass and Size. *Nat. Plants* 7 (1), 42–49. doi:10.1038/s41477-020-00815-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zinovyev, Sadovsky, Calzone, Fouché, Groeneveld, Chervov, Barillot and Gorban. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.