Technische Universität München
Fakultät für Mathematik

Mathematical modeling to dissect mechanisms of gene expression regulation

Lea Schuh, M.Sc.

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Christina Kuttler

Prüfer der Dissertation:

1. Prof. Dr. Dr. Fabian J. Theis

2. Prof. Dr. Arjun Raj (University of Pennsylvania)

3. Prof. Dr. Alexander Skupin (University of Luxembourg)

Die Dissertation wurde am 26.01.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 01.06.2022 angenommen.

Zusammenfassung

Genexpression bestimmt Funktion und Identität jeder Zelle. Denn obwohl alle Zellen unseres Körpers - von kleinen Veränderungen der DNA, entstanden durch Fehler während der Replikation oder Reparatur abgesehen - die gleiche genetische Information enthalten, führt jede Zelle ihre eigene und höchst spezialisierte Funktion aus. Wie genau eine Zelle ihre Genexpression reguliert, um zur richtigen Zeit die richtige Funktion auszuführen, und dies, obwohl sie ständig Genexpressionsrauschen und externen Signalen ausgesetzt ist, ist bis heute nur teilweise verstanden. Während neue experimentelle Methoden, wie die Einzelzell-RNA-Sequenzierung, große Datenmengen produzieren, die mittels maschinellem Lernen ausgewertet werden können, scheitern diese Ansätze für Datensätze von geringerer Größe. Werden etabliertere experimentelle Verfahren zur genaueren Untersuchung, beispielsweise über die Zeit, genutzt, die kleinere Datenmengen erzeugen, sind alternative Analysemethoden nötig. Falls ein ausreichendes Vorwissen der zugrundeliegenden Genregulation vorliegt, können mechanistische Modelle formuliert werden, die dieses Wissen einbeziehen. Dementsprechend sind mechanistische Modelle keine Patentlösung für jegliche Fragestellung, sondern müssen an jedes einzelne Problem und das entsprechende Vorwissen neu angepasst werden.

In der vorliegenden Arbeit wurden mechanistische Modelle aufgestellt, um genregulatorische Mechanismen unter Genexpressionsrauschen und externen Signalen zu verstehen. Zunächst wurde untersucht, ob Genexpressionsrauschen zu transienten und seltenen Veränderungen der Genexpression führen kann. Indem ein stochastisches Genexpressionsmodel entwickelt wurde, das sich auf Genexpressionsrauschen und etablierte Grundlagen der Genregulation stützt, konnte gezeigt werden, dass dieses einfache Model durchaus zu transienten und seltenen Genexpressionveränderungen führt. Stochastische Simulationen unseres Modells ähnelten Genexpressionsmuster vor-resistenter Melanomzellen einer Zellkultur. Als nächstes wurde untersucht, wie externe Signale wesentliche Bestandteile der Genexpressionsregulation darstellen können. Dafür ermittelten wir, ob die globale Genexpression während der Embryonalentwicklung des Frosches durch den Zellzyklus beeinflusst wird. Hierbei ist eine schnelle und ökonomische Genexpressionskontrolle notwendig, wenngleich die Zellen konstant dem Zellzyklus - in diesem Fall das sogenannte externe Signal - unterliegen. Durch die Formulierung einer Reihe von deterministischen mathematischen Modellen, die die Aktivität globaler Genexpression beschreiben, konnten wir den Zellzyklus als aktive und wichtige Komponente der Genexpressionsregulation identifizieren. Schließlich wurde untersucht, ob eine mehrfache Stimulation durch ein externes Signal zu einer adaptiven Genexpressionsregulation der Repression, dem Deaktivieren eines Gens, führt. Um dies zu beantworten, entwickelten wir deterministische Modelle, die die Genexpressionskinetik einzelner Hefezellen, die wechselnden Kohlenstoffquellen ausgesetzt waren, beschreiben und quantitativ vergleichen. Hiermit konnten wir zeigen, dass mehrfache externe Stimulation durchaus zu adaptiver Genexpressionsregulation der Repression führt.

Zusammengefasst haben wir stochastische und deterministische Modelle aufgestellt, die Genexpressionsrauschen und externe Signale berücksichtigen, um somit grundlegende genregulatorische Mechanismen zu untersuchen.

Abstract

The function and identity of a cell is determined by its gene expression levels. Hence, even though all cells in our body contain the same genetic information encoded in the DNA, that is up to small alterations due to errors in its repair or replication, each cell performs its own but highly specialized function. How exactly a cell regulates its correct gene expression at a given time to fulfill its specific function, although constantly subject to intrinsic noise and external cues, is only partially understood. While novel experimental techniques to study gene expression, such as single-cell RNA sequencing, produce data rich enough to be leveraged by machine learning and deep learning, these powerful approaches fail for small sample sizes. Hence, when using more established experimental techniques to investigate gene expression regulation on a smaller scale, for instance over time, other analysis approaches are needed. If prior knowledge of the underlying gene regulations is available, mechanistic models can be formulated to describe and study the phenomenon of interest. Accordingly, mechanistic models are not a one-size-fits-all solution, but need to be specifically tailored to the problem of interest and the available prior knowledge.

In this thesis, we developed and evaluated several mechanistic models to dissect mechanisms of gene expression regulation by investigating intrinsic noise and external cues. Firstly, we studied intrinsic noise in the context of transient and rare changes in gene expression. By developing a stochastic gene expression model comprising intrinsic noise and established principles of gene regulation, we were able to show that such a simplistic model allows for transient and rare changes in gene expression, similar to gene expression patterns observed in pre-resistant melanoma cells of a human skin cancer cell line. Next, we investigated how external cues can be an integral part of gene expression regulation. For this, we inquired whether global gene expression is actively regulated by the cell cycle during frog embryonic development. Here, a fast and economical control of gene expression is required, while cells are constantly exposed to rapid cell divisions, here the external cue. By formulating and evaluating a set of deterministic mathematical models describing global gene expression, we identified the cell cycle to be an active and important regulatory component of gene expression. Finally, we asked whether the repeated exposure to an external cue can lead to an adaptive gene expression regulation of repression, i.e., the deactivation of a gene. To answer this, we developed deterministic models describing and comparing the gene expression kinetics of single yeast cells, which were repeatedly exposed to carbon-source shifts. Thereby, we showed that a repeated external cue can indeed lead to an adapted gene expression regulation of repression.

In summary, we developed stochastic and deterministic models addressing the effects of different sources of intrinsic noise and external cues on gene expression regulation, which provide new insights into and allow for a more comprehensive picture of the mechanisms of gene expression regulation.

Acknowledgments

The list of people I want to thank and who have directly or indirectly contributed to this thesis is long.
Foremost, I want to thank Carsten Marr for believing in me when I didn't and taking me in as an intern in 2017. Everything else is a result or a consequence of those four weeks. Thank you for being the best supervisor and mentor I could have wished for during the past years. Your guidance and support exceeds anything a Ph.D. student could hope for. A huge thank you to Fabian Theis for making this Ph.D. possible and for being my university supervisor. Thank you for creating such an inclusive, open and exciting research environment. Further, I'd like to thank Robert Schneider for inviting me into his lab and allowing me to learn way more biology than I thought I would - ever. I feel really lucky to have had you as a collaboration partner and mentor. I'd further like to thank Arjun Raj for allowing me to write my master thesis in his lab - now you've been stuck with me for four years. Thank you for your input and continued support. A very special thank you to Yogesh Goyal. Thank you for your supervision, guidance and continued collaboration. But above all, thank you for being my mentor and friend. Thank you to my collaboration partners Poonam Bheda, Maria Bordukova, Ben Emert, Axel Imhof, Igor Kukhtevich, Carolin Loos, Daniil Pokrovsky, Ralph Rupp, Michael Saint-Antoine, Eric Sanford, Melanie Schulz, and Abhyudai Singh. I was really lucky to have had such fun and exciting collaborations. Thank you to Christina Kuttler - not only for taking over the chair of my defense committee but also for accompanying my Ph.D. journey. I was really lucky to have had your support and input throughout the past years. Another special thanks to Alexander Skupin for agreeing to be an examiner on my defense committee. Apart from that I would like to thank my lab members and colleagues, particularly: Ines Assum, Till Bartke, Lisa Bast, Volker Bergen, Ali Boushehri, Maren Büttner, Leander Dony, Olle Holmberg, Jakob van Höfer, Amelie Kraus, Marius Lange, Carlos Talavera Lopez, Saulius Lukauskas, Valerio Lupperger, Benedikt Mairhörmann, Marlet Morales, Francesco Padovani, Ario Sadafi, Yannik Schälte, Kurt Schmoller, Paul Stapor, Moritz Thomas, Dominik Waibel, Dantong Wang, and Johannes Wirth, and everyone I forgot to name here - you know who you are. A big thank you to the epigenetics community at the Helmholtz Zentrum München and the EpiCrossBorders Research School for all the additional scientific and non-scientific support. A special thank you to the Joachim Herz Foundation for your financial support and network throughout the past years. A special thanks to both the administration at the Computational Health Center and at TUM, namely Marianne Antunes, Sabine Kunz, Anna Sacher, and Lydia Weber. Thank you to my friends who have supported me in the past years, particularly Lisa Bernhardt, Marie Harder, Katharina Kempf, Elena Kern, Luca Schlegel, Annkathrin Swanson, and Fabian Wagner. I want to thank my parents, Sonja and Armin Schuh, and my sister, Anina Schuh, for their endless support. I couldn't have done this without you. And finally, I want to thank my biggest fan, Max Kisanyik. I do not even know where to start. I am so lucky to have you in my life.

List of contributed publications

This thesis is based on the following publications:

Core publications as main author

I) Schuh, L., Saint-Antoine, M., Sanford, E.M., Emert, B.L., Singh, A., Marr, C., Raj, A., and Goyal, Y. (2020). Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer. *Cell Systems*, *10*(4), 363-378. doi: https://doi.org/10.1016/j.cels.2020.03.004
(See also publication [1] in the bibliography.)

II) Schuh, L., Loos, C., Pokrovsky, D., Imhof, A., Rupp, R.A.W., and Marr, C. (2020). H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested Xenopus embryos. *Cell Systems*, *11*(6), 653-662. doi: https://doi.org/10.1016/j.cels.2020.11.003
(See also publication [2] in the bibliography.)

Further manuscript as main author currently under review

III) Schuh, L., Kukhtevich, I., Bheda, P., Schulz, M., Bordukova, M., Schneider, R., and Marr, C. (2022). Gal1 repression memory in budding yeast. *Submitted.*
(See also manuscript [3] in the bibliography.)

Other publications not included in this thesis

IV) Alabert, C., Loos, C., Voelker-Albert, M., Graziano, S., Forné, I., Reveron-Gomez, N., Schuh, L., Hasenauer, J., Marr, C., Imhof, A., and Groth, A. (2020). Domain model explains propagation dynamics and stability of histone H3K27 and H3K36 methylation landscapes. *Cell Reports*, *30*(4), 1223-1234. doi: https://doi.org/10.1016/j.celrep.2019.12.060
(See also publication [4] in the bibliography.)

V) Boe, R.H., Ayyappan, V., Schuh, L., Raj, A. (2021). Allelic correlation is a marker of trade-offs between barriers to transmission of expression variability and signal responsiveness in genetic networks. *bioRxiv 2021.11.26.470134*, doi: https://doi.org/10.1101/2021.11.26.470134
(See also manuscript [5] in the bibliography.)

I, Lea Schuh, am the main author of publications [1] and [2] and manuscript [3]. Detailed descriptions of my specific contributions for these two publications and the manuscript can be found as a summary in the introduction, section 1.6, and in full length in the appendices A.1, A.2 and B.1, respectively.

Contents

1. Introduction

Models are abstractions of reality and are used to explain the world [6]. As most phenomena are too complicated to be described in their entirety, a model is restricted to the most important components of a system, whereby the choice of interactions between components represented by the model strongly influences its performance. The common aphorism "All models are wrong, but some are useful", attributed to George Box, recognizes that all models are hence just better or worse approximations of reality [7]. Mathematical models describe these abstractions by mathematical formulations allowing for their analyses by means of theory and algorithms. One of the earliest translated mathematical models was developed by Ptolemy in 150 AD and describes the planetary motions of the solar system [8]. Still assuming that the earth is stationary and at the universe's center, Ptolemy described the motion of a planet by a circle rotating around a larger circle surrounding earth. Nonetheless, this model accurately predicted the planet's motion with respect to earth and was only replaced by the Copernican system and Johannes Keppler's laws of planetary motion in the 16th and 17th centuries. This is just one of many examples showing how a seemingly complex concept can have a simpler underlying explanation if seen the right way, where simplicity often yields truth. Mainly driven by astronomy and physics, a plethora of different classes of models emerged over the past centuries. Coupling established principles of dynamic mathematical modeling with systems biology, a rather new interdisciplinary field quantitatively analyzing complex biological systems, I described and studied models of gene expression regulation.

1.1. Gene expression regulation

The (genetic) information flow within a cell was first proposed by James D. Watson in 1965 [9]: in a first step, the information encoded by the gene, a basic functional unit of the deoxyribonucleic acid (DNA) composed of nucleotides, is transcribed to a complementary single-stranded mature ribonucleic acid (mRNA); in a second step the information encoded by the mRNA is translated into an amino-acid sequence, the protein (Figure 1.1).

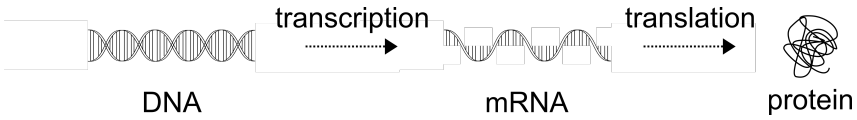


Figure 1.1: The genetic information encoded by the DNA is transcribed to mRNA, and is then translated to the functional gene product, called protein.

Today, we know that the true process of gene expression is more complex, depending on, e.g., the local DNA compaction and the availability and correct positioning of other proteins, such as regulatory signals and operating machineries [10–12]. Hence, the expression of a gene is strongly regulated by the protein levels of other genes, a dependency termed gene regulation, first described by Jacques L. Monod in a bacterial system in 1961 [13]. This means further, that the function and identity of a cell is determined and regulated by the expression levels of all its genes. Even though all cells in our body contain the same genetic information encoded in the

DNA, that is up to small alterations due to errors in its repair or replication, we are composed of a large variety of cells each performing its own function. How exactly a cell regulates its correct gene expression at a given time to fulfill its specific function, although constantly subject to intrinsic noise and external cues, is unknown.

1.2. Experimental techniques to study gene expression regulation

To experimentally study gene expression regulation, one is interested in quantifying the mRNA and/or protein levels of one or more genes. Recent advances in sequencing, i.e., the process of identifying the sequence of nucleotides, enables the measurement of single mRNA counts across all genes within a cell. This high-throughput technique is called single-cell RNA sequencing (scRNA-seq) and allows for the simultaneous study of gene expression in millions of cells [14]. For this, cells of a biological sample are isolated, by encapsulating them in individual droplets, and labeled with a unique barcode. This barcode is used to retrieve the cell-specific information after sequencing. More established, and preserving spatial information, is a technique referred to as single-molecule RNA fluorescence in situ hybridization, where individual mRNA molecules of a pre-defined gene can be detected in single, fixed cells [15]. Here, probes attached to fluorophores are hybridized to single mRNA molecules, which are then visible as single bright spots under the microscope. Similarly to smRNA FISH, gene expression reporters, such as the green fluorescent protein (GFP), are a state-of-the-art method to measure single-cell protein levels [16]. Here, the GFP sequence is inserted into the gene. Upon transcription and translation, the gene-GFP fusion is expressed and can be quantified by imaging, where the GFP fluorescence intensities are assumed to be informative about the expression levels of the gene of interest. In contrast to scRNA-seq and smRNA FISH, the gene-GFP fusion can be used to track single-cell gene expression over time. Moreover, for packaging and organizing the three-meter long DNA into the nucleus of a cell, which is not larger than a few micrometers, the DNA is wrapped around histone proteins [17]. In the early 1960s, Vincent Allfrey identified that histone proteins do not only have a structural function, but can also be chemically modified, altering the local DNA compaction and thereby facilitating or inhibiting gene expression [18, 19]. Hence, histone modifications allow for a dynamic regulation of gene expression. The relative abundance of specific histone modifications are nowadays used as a proxy for global gene activity across cell populations and can be determined by mass spectrometry [20].

1.3. Mathematical models of gene expression regulation

Although experimental techniques are invaluable for measuring gene expression levels, they do not offer insights into the mechanisms of gene expression regulation. While the rich information of data acquired by scRNA-seq can be leveraged by novel mathematical and computational methods from the machine learning and deep learning field [21–23], where little to no prior knowledge is required, these powerful methods fail for small sample sizes. Hence, when using experimental techniques to investigate gene expression regulation on a smaller scale, alternative approaches regarding analysis are needed. By using prior knowledge on the underlying gene regulations, mechanistic models can be formulated to describe and study the phenomenon of interest. Accordingly, mechanistic models are not a one-size-fits-all solution, but need to be

specifically tailored to the problem of interest. In comparison to machine and deep learning models, mechanistic models provide an easily interpretable output. Hence, mechanistic models were the first models developed to describe the dynamical processes of gene expression and are still commonly applied today.

Brian C. Goodwin introduced the first mathematical models of gene expression in 1965 shortly after its principles were discovered [24]. Goodwin's models describe the mRNA and protein concentrations of (i) a self-repressing gene and (ii) a pair of mutually repressing genes by ordinary differential equations [25]. Using mathematical theory and an analogue computer for simulations, he discovered that the expression of a single gene could be governed by an undamped oscillatory behavior. John S. Griffith refuted Goodwin's hypothesis only three years later, upon which Goodwin claimed the simulations to have been erroneous [26]. Nonetheless, Goodwin discussed the results of his modeling in light of adaptive systems, i.e., organisms with circadian organization, highlighting the strong dependence between gene expression regulation and external cues.

When considering single cells, Aaron Novick and Milton Weiner described the expression of a certain protein already in 1957 as 'an all or none-phenomenon', where expression seemed to be 'the result of a random single event' [27]. Early studies of the variability of gene expression between cells were limited due to the lack of reliable single-cell gene expression reporters [28]. Only in 1990 did Ko et al. develop an assay to identify single-cell gene expression levels and reported a then surprisingly large variability, which could not be explained by deterministic gene expression models [29]. Today, it is well known that genetically identical cells exposed to the same environmental conditions can demonstrate large fluctuations in their gene expressions, a phenomenon termed cell-to-cell variability [30]: gene expression involves a succession of probabilistic events, where a gene is present in just one or a few copies per cell and is activated (inactivated) by random associations (dissociations) leading to noise intrinsic to the process of gene expression itself. To account for intrinsic noise and describe the stochastic nature of these reactions, Minoru S. H. Ko developed the first stochastic model for gene induction, called the telegraph model [31]. The gene is assumed to transition between an active state and an inactive state, in which mRNA is and is not transcribed, respectively. Computer simulations of the telegraph model demonstrated heterogeneous levels of gene induction, comparable to the experimentally observed intrinsic noise. Raj et al. have verified that transcription occurs in short and intense bursts, a phenomenon today termed transcriptional bursting [28, 32].

With the formation of the research field of systems biology in the beginning of the 2000s [33], the focus of both deterministic and stochastic studies shifted to systems of coupled gene expression regulations summarized by gene regulatory networks. Within systems biology, deterministic and stochastic models of gene expression have been widely developed to explain different biological phenomena ranging from a more detailed understanding of transcription [34, 35] and translation [36, 37], to embryonic patterning [38–41], pathophysiology [42, 43], circadian clocks [44], to stress response [45] and adaptation [46].

However, how intrinsic noise and external cues are integrated into the mechanisms of gene expression regulation is still an open question for many biological and biomedical systems. In this thesis, we couple experimental data analysis, computational simulations, and mathematical modeling to gain a more informed and quantitative representation of gene expression regulation.

1.4. Research questions

The main goal of this thesis is to dissect mechanisms of gene expression regulation by investigating the effects of intrinsic noise and external cues on expression dynamics using mathematical models. More specifically, we asked the following questions:

I) Is intrinsic noise sufficient to lead to transient and rare changes in gene expression?

II) Can cell cycle as a recurrent external cue be an integral part of gene expression regulation?

III) Can a repeated external cue lead to an adaptive gene expression response of repression?

1.5. Summary of results

This thesis contains two core publications and one manuscript under review, in which we explored the different effects of intrinsic noise and external cues on gene expression regulation as listed in the previous section. Doing so appropriately, we explored several model systems from human melanoma cells over frog to yeast. In Schuh et al. [1] we explored research question I): Rare and transient gene expression patterns have previously been observed in pre-resistant melanoma (skin cancer) cells, where cells in the rare gene expression state at time of drug administration are linked to survival [47]. To understand whether intrinsic noise is sufficient to lead to such rare and transient gene expression states, we developed a stochastic mathematical model and compared our simulations to experimental data from smRNA FISH. In Schuh et al. [2] we explored research question II): During development, a fast and economical control of gene expression is required, while cells are constantly exposed to rapid cell divisions - here, the cell cycle represents the external cue. By formulating and evaluating a set of deterministic mathematical models on mass spectrometry data of a specific histone modification, which acts as a proxy for global gene expression, of frog embryos, we investigated whether global gene expression is actively regulated by the cell cycle. In other words, we studied whether the cell cycle is an integral part of gene expression regulation during frog development. Finally, in Schuh et al. [3], we explored research question III): To study whether cells repeatedly exposed to an external cue - here, repeated carbon source shifts - show adaptive gene expression kinetics in repression, we developed a deterministic model describing the gene repression kinetics of single yeast cells. Using our model we compared the single-cell gene repression kinetics of yeast between repeated carbon source shifts, where single-cell gene expression kinetics were experimentally monitored by a gene-GFP fusion. A detailed summary of the results of these studies can be found below.

Core publications as main author

- [1] in Appendix A.1: *Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer*
  Melanoma cells have been linked to a coordinated and extremely high gene expression in a number of specific marker genes, arising and disappearing over time [47]. Only cells with coordinated high expression at the timing of drug administration are thought to be pre-resistant, surviving drug therapy and acquiring stable resistance. To identify the origins of these gene expression patterns, we developed a mathematical model consisting

of transcriptional bursting, gene regulations and stochasticity in gene expression. We found that given specific parameter sets and gene regulatory networks, our simulated gene expressions qualitatively recapitulate the patterns observed in pre-resistant melanoma cells. These states occur more frequently in gene regulatory networks with low connectivity. Furthermore, we found that bursting and regulatory parameters in the model are determinant of the number of simulations with desired gene expression patterns. Using the simulations, we further studied the initiation, maintenance and termination of these expression states and found that a long transcriptional burst initiates, an increase in burst frequency maintains and a random process terminates these expression states. Finally, we used our mathematical model to hypothesize that stable resistance after drug administration may be initiated by more gene regulation, which we confirmed experimentally. Together, we demonstrate that intrinsic noise and established principles of gene regulation are sufficient to describe transient and rare gene expression, qualitatively similar to the gene expression patterns as observed in pre-resistant melanoma cells.

- [2] in Appendix A.2: *H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested Xenopus embryos*
  Gene expression depends on the local compaction of DNA. The openness of DNA is regulated, among other mechanisms, by the placement of methyl groups on the tails of histone proteins. However, cells divide rapidly during development and with each cell cycle unmodified histones get incorporated into the DNA, leading to an overall dilution of histone modifications. To study whether the cell cycle actively regulates the histone modification landscape, and hence global gene expression, during development, we compared the methylation kinetics of a specific histone modification site, H4K20, between cycling and cell-cycle-arrested *Xenopus laevis* (frog) embryos. To identify whether population-specific parameters are required to explain the H4K20 methylation kinetics between cycling and cell-cycle-arrested embryos, we formulated a set of plausible models, containing joint and population-specific parameters. Performing multi-start maximum likelihood optimization and model selection on all models, we found that population-specific parameters were indeed required to explain the H4K20 methylation kinetics of both embryo populations. This suggests that the cell cycle plays an active role in shaping the H4K20 methylation kinetics and, hence, global gene expression regulation. Moreover, we identified that demethylation, the active removal of methyl groups, is essential for regulating H4K20 methylation kinetics in cell-cycle-arrested embryos, while it is likely dispensable in cycling embryos. This indicates that cell-cycle-mediated dilution is an essential regulatory component for shaping the H4K20 methylation landscape during early development. Together, we demonstrate that an external cue, such as the cell cycle during frog development, can be an integral part of global gene expression regulation.

Further manuscript as main author currently under review

- [3] in Appendix B.1: *Gal1 repression memory in budding yeast*
  When exposed to repeated external cues, cells can 'remember' their initial activating

response, leading to an adapted reaction in gene expression upon re-stimulation, a phenomenon termed transcriptional memory. Most studies on transcriptional memory have focused on gene induction, the activation of gene expression, although gene repression, the deactivation of gene expression, plays a similarly central role in gene expression regulation. This opens the important question as to whether memory also exists for repression. To address this, we measured the gene expression of gene *Galactokinase 1* (Gal1) through repeated carbon source shifts in single budding yeast cells by GFP-reporters. As Gal1 is sensitive to changes in the carbon source available to the cell, Gal1 gene expression showed induction and repression cycles. To deconvolute the kinetics of dilution via budding, the developing of a new cell, and repression, we first generated a method to compensate for dilution effects. Then, we formulated two mathematical models based on ordinary differential equations to appropriately describe the repression kinetics of repressor and non-repressor cells, i.e., cells with and without repression kinetics. By performing multi-start maximum likelihood optimization on single-cell repression traces we inferred the single-cell parameters per repression. Furthermore, we applied model selection to quantitatively discriminate between repressor and non-repressor cells. Using statistical testing, we identified the repression delay, i.e., the time between the beginning of the repression cycle and actual repression initiation, as a model parameter significantly decreased in repressor cells of repeated repression. The altered repression kinetics between consecutive repression cycles suggests that there is repression memory. Moreover, using the same modeling approach, we identified a gain-of-repression memory mutant, *elp6*Δ, exhibiting a stronger fold change between the repression delays between first and second repression. Together, we show that a repeated external cue, such as repeated carbon source shifts, can lead to adaptive gene expression regulation of repression.

1.6. Summary of individual contributions

To outline and emphasize my individual contributions of the two core publications and the manuscript contained in this thesis, I here provide a summary of the statements of individual contributions. More detailed and full-length descriptions of my specific contributions per publication/manuscript can be found in the appendices A.1, A.2 and B.1.

Core publications as main author

- [1] in Appendix A.1: *Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer*
  In my master thesis, I formulated an initial mathematical model to produce simulations recapitulating the observed gene expression patterns of pre-resistant melanoma cells [48]. During my Ph.D., I refined and extended the mathematical model, validated its generalizability with respect to fixed model parameters, and performed a large-scale simulation study to systemically screen and analyze the gene expression patterns of a self- and pre-defined set of gene regulatory networks and parameter sets. Merely, the validation of our hypothesis regarding stable resistance formation was experimentally validated by our collaboration partners, Michael Saint-Antoine and Abhyudai Singh. Otherwise, I was responsible for

the entire computational methodology and formal analysis. The original manuscript draft excluding the methods section was written by Yogesh Goyal and reviewed as well as edited by me and others. The methods section *STAR METHODS* was written by me. I generated all the figures. The full-length statement of my individual contribution to this publication can be found in A.1.

- [2] in Appendix A.2: *H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested Xenopus embryos*
  I formulated a mathematical model describing the histone methylation kinetics for both cycling and cell-cycle-arrested Xenopus embryos, identified a subset of plausible models by optimizing the mathematical formulation based on ordinary differential equations on a time-series mass spectrometry data set of H4K20 methylation [49], and identified joint and cycling-specific rate constants. I was responsible for the entire computational methodology, formal analysis and the writing of the manuscript. Merely the subsection *Experimental model and subject details* in *STAR METHODS*, describing the experimental work, was written by my co-author Daniil Pokrovsky. The full-length statement of my individual contribution to this publication can be found in A.2.

Further manuscript as main author currently under review

- [3] in Appendix B.1: *Gal1 repression memory in budding yeast*
  I formulated the mathematical models for the repressor and non-repressor cells, developed a method to compensate for dilution effects in the Gal1 kinetics and performed parameter estimation and model selection on the single-cell Gal1 traces. Moreover, I performed the statistical analysis on the estimated parameters of the repressor cells and applied the same modeling approach to the Gal1 repression kinetics of the *elp6*Δ mutant. I was responsible for the entire computational methodology, formal analysis and writing of the manuscript. The full-length statement of my individual contribution to this manuscript can be found in B.1.

1.7. Outline

In the next chapters, the main mathematical concepts of this thesis and the included publications are presented and integrated into the existing literature and broader scientific context. More specifically, in Chapter 2, we provide a short introduction to graph theory which was used to describe the inter-gene regulations summarized by gene regulatory networks. In Chapter 3, we discuss the mathematical description of chemical reaction networks, derive the chemical master equation and introduce Gillespie's stochastic simulation algorithm. As an application, we briefly present the stochastic gene expression model developed in Schuh et al. [1] used to study the effects of intrinsic noise on transient and rare gene expression patterns. In Chapter 4, we derive the deterministic description of the evolution of the mean of a chemical reaction network and introduce the deterministic model formulation. In Chapters 5 and 6, we discuss methods of parameter estimation and model selection as applied throughout the aforementioned publications [1, 2] and manuscript [3]. Finally, we briefly present the deterministic models to

study the effects of external cues on gene expression regulation as developed in Schuh et al. [2] and Schuh et al. [3] as applications.
The core publications [1] and [2] are included in Appendix A. Appendix B contains the manuscript under review [3]. Every publication and manuscript is preceded by a summary, a full-length statement highlighting my individual contribution, and, if needed, a permission to include the publication in this thesis given by the respective journal.

2. Graph theory

The paper on the Seven Bridges of Königsberg by Leonhard Euler published in 1736 is considered to mark the beginning of modern graph theory [50, 51]. Since then, graph theory has witnessed an unprecedented growth, today being applied in many fields including biology, electrical engineering, computer science and operations research [52–55].
A graph is a mathematical structure of pairwise relations (edges) between discrete objects (vertices). In this thesis, we use graph theory to describe the relationships between genes, where the vertices in a graph represent genes, and the edges represent gene regulations. These graphs are termed gene regulatory networks [56]. In Schuh et al. [1], we compared the simulations of gene expression patterns for which the underlying gene regulatory networks varied in the number of genes and their regulations. To define a suitable set of small gene regulatory networks for which we simulated the gene expression patterns, we made use of graph theoretical properties as introduced in this chapter, which is mainly adapted from [57–59].

2.1. Undirected graphs

In this section, we present a collection of definitions introducing graph theoretical concepts as applied in Schuh et al. [1].

Definition 2.1. *An undirected graph is an ordered pair $G = (\mathcal{V}(G), \mathcal{E}(G))$, where $\mathcal{V}(G) = \mathcal{V} \neq \emptyset$ is a finite non-empty set and $\mathcal{E}(G) = \mathcal{E}$ with $\mathcal{E} \subseteq \{\{v_i, v_j\} \subseteq \mathcal{V}^2 : v_i, v_j \in \mathcal{V}, v_i \neq v_j\}$ unordered pairs of elements of $\mathcal{V}$. Elements of $\mathcal{V}$ are called the vertices (or nodes) and elements of $\mathcal{E}$ are called the edges of G.*

Notation. For simplicity, we write $G = (\mathcal{V}, \mathcal{E})$ if it is clear that $\mathcal{V}$ and $\mathcal{E}$ belong to graph G.

Definition 2.2. *The order of an undirected graph $G = (\mathcal{V}, \mathcal{E})$ is defined as the number of vertices $\mathcal{V}$ of G, where* $\text{order}(G) = |\mathcal{V}|$.

Definition 2.3. *Two vertices $v_i, v_j \in \mathcal{V}$ are called adjacent if and only if $\{v_i, v_j\} \in \mathcal{E}$.*

Definition 2.4. *The neighborhood of vertex $v_i \in \mathcal{V}$ is defined by $\mathcal{N}(v_i) = \{v_j \in \mathcal{V} : \{v_i, v_j\} \in \mathcal{E}\}$.*

Definition 2.5. *The degree of vertex $v_i \in \mathcal{V}$ is defined by* $\deg(v_i) = |\{e \in \mathcal{E} : v_i \in e\}| = |\mathcal{N}(v_i)|$.

Definition 2.6. *A v_1, v_{n+1}-walk in an undirected graph $G = (\mathcal{V}, \mathcal{E})$ is a sequence $W = \langle v_1, e_1, v_2, e_2, ..., e_i, v_{i+1}, ..., e_n, v_{n+1} \rangle$ of vertices $v_1, ..., v_{n+1} \in \mathcal{V}$ and edges $e_1, ..., e_n \in \mathcal{E}$ such that $\forall e_j = \{v_j, v_{j+1}\} \in \mathcal{E}$.*

Definition 2.7. *An undirected graph $G = (\mathcal{V}, \mathcal{E})$ is called connected if $\forall v_i, v_j \in \mathcal{V}$ $\exists$ a v_i, v_j-walk.*

The regulatory properties of an undirected graph can be summarized by a matrix as defined below.

Definition 2.8. *For an undirected graph $G = (\mathcal{V}, \mathcal{E})$ of* $\text{order}(G) = |\mathcal{V}|$*, the adjacency matrix $A \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$, is a matrix whose rows and columns are indexed by the same orderings of $\mathcal{V}$ such that*

$$A_{ij} = \begin{cases} 1, & \text{if } \{v_i, v_j\} \in \mathcal{E}, \\ 0, & \text{else.} \end{cases} \tag{2.1}$$

Hence, the adjacency matrix of an undirected graph is symmetric.

Example. Let us consider an undirected graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, v_3, v_4, v_5\}$ and $\mathcal{E} = \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_1, v_4\}\}$ as depicted in Figure 2.1.

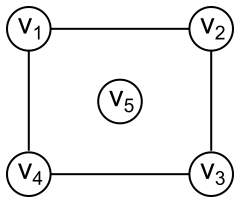


Figure 2.1: Exemplary undirected graph G.

- The order of G is given by $\text{order}(G) = 5$.
- Vertices v_1 and v_2 are adjacent.
- The neighborhood of v_1 is given by $\mathcal{N}(v_1) = \{v_2, v_4\}$.
- The degree of v_1 is given by $\deg(v_1) = |\mathcal{N}(v_1)| = 2$.
- $W = \langle v_1, \{v_1, v_2\}, v_2, \{v_2, v_3\}, v_3, \{v_3, v_4\}, v_4 \rangle$ is a walk connecting vertices v_1 and v_4.
- G is not connected as there exists no walk from v_1 to v_5.
- The adjacency matrix of G is given given by

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

2.2. Digraphs

Gene regulation is directed, i.e., gene X regulates gene Y, or gene Y regulates gene X. Furthermore, genes are able to auto-regulate their own gene expressions, i.e., gene X regulates gene X. As the previous definitions regarding undirected graphs are not sufficient to describe these gene regulatory network properties, we further define directed graphs.

Definition 2.9. *A (non-simple) digraph is an ordered pair $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} \neq \emptyset$ is a finite non-empty set and $\mathcal{E}$ with $\mathcal{E} \subseteq \{(v_i, v_j) \subseteq \mathcal{V}^2 : v_i, v_j \in \mathcal{V}\}$ a set of ordered pairs of elements of $\mathcal{V}$.*

A (non-simple) digraph allows for self-loops e.g., edge $e = (v_i, v_i)$, for $v_i \in \mathcal{V}$.
Accordingly, we reformulate the definitions regarding order, neighborhood and degree to account for digraphs.

Definition 2.10. *The order of a digraph $G = (\mathcal{V}, \mathcal{E})$ is defined as the number of vertices $\mathcal{V}$ of G, where* $\text{order}(G) = |\mathcal{V}|$.

Definition 2.11. *The out-neighborhood of vertex $v_i \in \mathcal{V}$ is defined by $\mathcal{N}_{out}(v_i) = \{v_j \in \mathcal{V} : (v_i, v_j) \in \mathcal{E}\}$. The in-neighborhood of vertex $v_i \in \mathcal{V}$ is defined by $\mathcal{N}_{in}(v_i) = \{v_j \in \mathcal{V} : (v_j, v_i) \in \mathcal{E}\}$.*

Definition 2.12. *The out-degree of vertex $v_i \in \mathcal{V}$ is defined by* $\deg_{\text{out}}(v_i) = |\mathcal{N}_{out}(v_i)|$ *and the in-degree of vertex $v_i \in \mathcal{V}$ is defined by* $\deg_{\text{in}}(v_i) = |\mathcal{N}_{in}(v_i)|$.

Definition 2.13. *A digraph $G = (\mathcal{V}, \mathcal{E})$ is called weakly connected if the underlying undirected graph is connected.*

Similar to undirected graphs, the regulatory properties of digraphs can be summarized by a matrix as defined below.

Definition 2.14. *For a digraph $G = (\mathcal{V}, \mathcal{E})$ of* $\text{order}(G) = |\mathcal{V}|$*, the adjacency matrix $A \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$, is a matrix whose rows and columns are indexed by the same orderings of $\mathcal{V}$ such that*

$$A_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in \mathcal{E}, \\ 0, & \text{else.} \end{cases} \tag{2.2}$$

2.3. Additional structural graph properties

Definition 2.15. *Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2)$ be two digraphs. G_1 and G_2 are called isomorphic if and only if there exists a bijection $f : \mathcal{V}_1 \to \mathcal{V}_2$ such that $\forall v_i, v_j \in \mathcal{V}_1 : (v_i, v_j) \in \mathcal{E}_1 \Leftrightarrow (f(v_i), f(v_j)) \in \mathcal{E}_2$. The bijection f is called an isomorphism of G_1 onto G_2.*

Definition 2.16. *A digraph $G = (\mathcal{V}, \mathcal{E})$ of* $\text{order}(G) = |\mathcal{V}|$ *is called symmetric if the rows of its adjacency matrix A are cyclic permutations to the right with offset one of each other. The adjacency matrix A of symmetric digraph G is then given by*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1|\mathcal{V}|} \\ a_{1|\mathcal{V}|} & a_{11} & a_{12} & \dots & a_{1(|\mathcal{V}|-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{12} & a_{13} & a_{14} & \dots & a_{11} \end{pmatrix}. \tag{2.3}$$

Each vertex in a symmetric digraph has the same number of in- and outgoing edges and self-loops. Hence, the in-degrees and out-degrees of all vertices in a symmetric digraph are equal. Moreover, in a symmetric digraph, all vertices are contextually identically embedded in the graph.

Example. Let us consider a digraph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, v_3, v_4\}$ and

$$\mathcal{E} = \{(v_1, v_1), (v_1, v_2), (v_1, v_3), (v_1, v_4), (v_2, v_1), (v_3, v_1), (v_4, v_2), (v_4, v_3)\}$$

and digraph $H = (\mathcal{W}, \mathcal{F})$, where $\mathcal{W} = \{w_1, w_2, w_3, w_4\}$ and

$$\mathcal{F} = \{(w_1, w_1), (w_1, w_2), (w_1, w_3), (w_1, w_4), (w_2, w_1), (w_3, w_2), (w_3, w_4), (w_4, w_1)\}$$

as depicted in Figure 2.2.

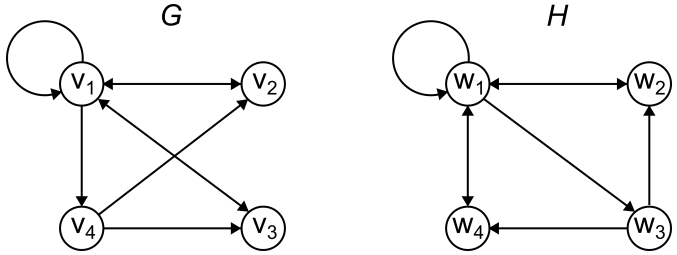


Figure 2.2: Exemplary directed digraphs G (left) and H (right).

- The orders of digraphs G and H are given by $\text{order}(G) = \text{order}(H) = 4$.
- The out-neighborhood of v_1 is given by $\mathcal{N}_{out}(v_1) = \{v_1, v_2, v_3, v_4\}$. The in-neighborhood of v_1 is given by $\mathcal{N}_{in}(v_1) = \{v_1, v_2, v_3\}$.
- The out-degree of v_1 is given by $\deg_{\text{out}}(v_1) = |\mathcal{N}_{out}(v_1)| = 4$. The in-degree of v_1 is given by $\deg_{\text{in}}(v_1) = |\mathcal{N}_{in}(v_1)| = 3$.
- G and H are weakly connected as the underlying undirected graph is connected.
- The adjacency matrix of digraph G is given by
$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$
- G and H are isomorphic under the bijection $f : v_1 \mapsto w_1, f : v_2 \mapsto w_2, f : v_3 \mapsto w_4, f : v_4 \mapsto w_3$.
- G and H are not symmetric digraphs as the rows of their adjacency matrices are not cyclic permutations to the right with offset one of each other.

3. Stochastic gene expression models

In Schuh et al. [1], we developed a mathematical model to study intrinsic noise and the formation of rare and transient gene expression states. For this, we formulated a model comprising gene interactions, transcriptional bursting and stochasticity and compared its simulations with rare and transient gene expression patterns observed in pre-resistant melanoma cells. To model stochasticity in gene expression appropriately, we used concepts originally introduced for modeling chemical reactions [60–63].

A chemical reaction is a conversion process during which one or more atoms, molecules or compounds are converted to one or more different substances. To study chemical reactions under a wide range and highly controlled conditions, mathematical and computational models were established. However, only in the late 1980s were mathematical models used to describe the kinetics of chemical reactions, a field today termed chemical reaction network theory [64].

Biological processes such as transcription, translation, and mRNA/protein degradation can be viewed as chemical reactions and studied by means of chemical reaction network theory. Stochastic models derived from chemical reaction network theory have been successfully and widely developed to study gene expression regulation, ranging from more theoretical studies of mRNA steady-state distributions assuming transcriptional bursting [65], to investigating the dynamics and attractor states of gene regulatory network modules [66], to describing lineage decision-making in single-cells, thereby, challenging existing hypotheses regarding the gene regulatory networks underlying such decisions [67]. This chapter is mainly based on [64, 68, 69], if not stated otherwise.

3.1. Mathematical formulation

A chemical reaction network describes the possible reactions within a reaction system. Here, we collect the most essential definitions in that context.

Definition 3.1. *A chemical reaction network is defined by a set of chemical species* $\mathcal{C} = \{C_1, ..., C_n\}$*, a set of chemical reactions* $\mathcal{R} = \{R_1, ..., R_m\}$ *and stoichiometric coefficients* $\nu_{ij} \in \mathbb{N}_0$ *and* $\eta_{ij} \in \mathbb{N}_0$ *such that reaction* $R_j \in \mathcal{R}$ *is described by*

$$\sum_{i=1}^{n} \nu_{ij} C_i \xrightarrow{k_j} \sum_{i=1}^{n} \eta_{ij} C_i, \tag{3.1}$$

where $k_j \in \mathbb{R}^+$ *is the reaction rate constant. Chemical species with* $\nu_{ij} \neq 0$ *are called educts and chemical species with* $\eta_{ij} \neq 0$ *are called products of reaction* R_j*.*

Definition 3.2. *The stoichiometric matrix* $S \in \mathbb{N}_0^{n \times m}$ *summarizes the net gains and losses of each chemical species per reaction.* S *is defined by the stoichiometric coefficients,* $\nu_{ij}, \eta_{ij} \in \mathbb{N}_0$*, such that* $S_{ij} = \eta_{ij} - \nu_{ij}$*.*

Definition 3.3. *The order of reaction* $R_j \in \mathcal{R}$ *is defined as*

$$o_j = \sum_{i=1}^{n} \nu_{ij}. \tag{3.2}$$

Definition 3.4. *The state of the chemical reaction network, i.e., the count per chemical species*

at time t, is described by the state vector $\mathbf{X}(t) \in \mathbb{N}_0^n$, where $\mathbf{X}(t) = [X_1(t), ..., X_n(t)]$ and n is the number of chemical species [70].

Example. Let us consider a common stochastic model describing the induction of a single gene, called a leaky telegraph model, first introduced by Minoru S. H. Ko in 1991 (Figure 3.1) [31]. DNA is either in an inactive ($\mathrm{DNA_{off}}$) or active ($\mathrm{DNA_{on}}$) state, where r_{on} and r_{off} the activating and deactivating rate constants. mRNA is transcribed with rate constants r_{prod} and dr_{prod} in the inactive and active DNA states, respectively, where $d > 1$, to ensure faster transcription in $\mathrm{DNA_{on}}$. The mRNA transcription in $\mathrm{DNA_{off}}$ is called leaky transcription. mRNA degrades with rate constant r_{deg}.

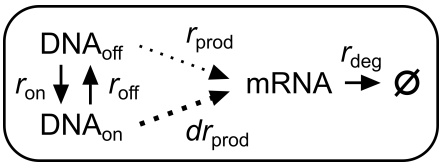


Figure 3.1: Schematic of the leaky telegraph model. The dotted arrows represent reactions without conversions, i.e., DNA is a template from which mRNA is transcribed, which maintains its state upon transcription. Adapted from [1].

The leaky telegraph model is described by chemical species $\mathcal{C} = \{\mathrm{DNA_{off}}, \mathrm{DNA_{on}}, \mathrm{mRNA}\}$ and reactions

$$\begin{aligned}
\text{DNA activation} \quad & R_1\text{:} \quad \mathrm{DNA_{off}} \xrightarrow{r_{\mathrm{on}}} \mathrm{DNA_{on}} \\
\text{DNA deactivation} \quad & R_2\text{:} \quad \mathrm{DNA_{on}} \xrightarrow{r_{\mathrm{off}}} \mathrm{DNA_{off}} \\
\text{leaky transcription} \quad & R_3\text{:} \quad \mathrm{DNA_{off}} \xrightarrow{r_{\mathrm{prod}}} \mathrm{DNA_{off}} + \mathrm{mRNA} \\
\text{transcription} \quad & R_4\text{:} \quad \mathrm{DNA_{on}} \xrightarrow{dr_{\mathrm{prod}}} \mathrm{DNA_{on}} + \mathrm{mRNA} \\
\text{degradation} \quad & R_5\text{:} \quad \mathrm{mRNA} \xrightarrow{r_{\mathrm{deg}}} \emptyset.
\end{aligned}$$

The state of the system is described by $\mathbf{X}(t) = [X_{\mathrm{DNA_{off}}}(t), X_{\mathrm{DNA_{on}}}(t), X_{\mathrm{mRNA}}(t)]$, with $X_{\mathrm{DNA_{off}}}(t)$, $X_{\mathrm{DNA_{on}}}(t) \in \{0, 1\}$, where 1 (0) denotes that the system is (not) in the given DNA state and $X_{\mathrm{mRNA}}(t) \in \mathbb{R}_{\geq 0}$, $\forall t \in \mathbb{R}^+$. The biological DNA states are mutually exclusive, i.e., $X_{\mathrm{DNA_{off}}}(t) + X_{\mathrm{DNA_{on}}}(t) = 1$, $\forall t \in \mathbb{R}^+$.

3.2. Law of mass action

First introduced by Peter Waage and Cato Guldberg in 1864, the law of mass action describes chemical reactions of solutions in dynamic equilibrium [71]. The general idea is that the more educts are present in a reaction system the more occurrences of the corresponding reaction there will be [68].

Definition 3.5. *Let us assume an idealized, well mixed system of constant volume which is in thermal equilibrium and kept at a constant temperature. Then the law of mass action states that the probability of reaction $R_j \in \mathcal{R}$ to occur in an infinitesimal time interval $[t, t + dt)$ is given by*

$$\mathbb{P}(R_j | \mathbf{X}(t)) = a_j(\mathbf{X}(t)) = c_j \prod_{i=1}^{n} \binom{X_i(t)}{\nu_{ij}}, \tag{3.3}$$

where $a_j(\mathbf{X}(t)) \in \mathbb{R}^+$ is called the reaction propensity and c_j a constant [70, 72].

Here, c_j describes the probability of a successful reaction to occur in $[t, t+dt)$ and the second term summarizes the potential number of educt collisions which could lead to this reaction. In general, $c_j \neq k_j$, where k_j is the reaction rate constant as defined in Definition 3.1.

Example. Reaction propensities for four so-called elementary reactions are stated below [70].

reaction	propensity	
$\emptyset \xrightarrow{k}$ products	k	
$C_i \xrightarrow{k}$ products	$kX_i(t)$	
$C_i + C_j \xrightarrow{k}$ products	$kX_i(t)X_j(t)$	$(i \neq j)$
$2C_i \xrightarrow{k}$ products	$kX_i(t)(X_i(t)-1)/2$	

Reactions containing more than two educts can be understood as multiple two-educt reactions and are hence not specified any further.

Example. For the leaky telegraph model the reaction propensities corresponding to reactions R_1, R_2, R_3, R_4 and R_5 are given by

$$\begin{aligned}
a_1(\mathbf{X}(t)) &= r_{\text{on}} X_{\text{DNA}_{\text{off}}}(t) \\
a_2(\mathbf{X}(t)) &= r_{\text{off}} X_{\text{DNA}_{\text{on}}}(t) \\
a_3(\mathbf{X}(t)) &= r_{\text{prod}} X_{\text{DNA}_{\text{off}}}(t) \\
a_4(\mathbf{X}(t)) &= dr_{\text{prod}} X_{\text{DNA}_{\text{on}}}(t) \\
a_5(\mathbf{X}(t)) &= r_{\text{deg}} X_{\text{mRNA}}(t).
\end{aligned}$$

3.3. Chemical master equation

The chemical master equation was first derived by Daniel T. Gillespie in 1992 and is a system of ordinary differential equations (ODEs) describing the time evolution of the probability distribution of the state vector $\mathbf{X}(t)$ of a chemical reaction network [72]. The chemical master equation is derived by taking together the following three theorems:

Theorem 3.1. *Let $\mathbf{X}(t) \in \Gamma \subseteq \mathbb{N}_0^n$ be an element of the state space Γ for time t. If $\mathbf{X}(t+dt) = \mathbf{X}(t) - \mathbf{s_j}$, where $\mathbf{s_j}$ the j^{th} column vector of stoichiometric matrix S, then the probability that exactly one reaction $R_j \in \mathcal{R}$ will occur in time interval [t,t+dt) is equal to*

$$a_j(\mathbf{X}(t) - \mathbf{s_j})dt + o(dt), \tag{3.4}$$

where a_j the reaction propensity of reaction R_j [72].

Theorem 3.2. *If $\mathbf{X}(t+dt) = \mathbf{X}(t)$, then the probability that no reaction will occur in time interval $[t, t+dt)$ is equal to*

$$\prod_{j=1}^{m}(1 - a_j(\mathbf{X}(t))dt) + o(dt) = 1 - \sum_{j=1}^{m} a_j(\mathbf{X}(t))dt + o(dt). \tag{3.5}$$

Adapted from [72].

Theorem 3.3. *The probability of more than one reaction occurring in time interval $[t, t+dt)$ is $o(dt)$ [72].*

For the proofs of the previous theorems see [72]. Together, these theorems express the probability

to end in state $\mathbf{X}$ at time $t + dt$ denoted by $p(\mathbf{X}, t + dt)$. Hence,

$$p(\mathbf{X}, t + dt) = p(\mathbf{X}, t)(1 - \sum_{j=1}^{m} a_j(\mathbf{X})dt + o(dt)) + \sum_{j=1}^{m} p(\mathbf{X} - \mathbf{s_j}, t)\left(a_j(\mathbf{X} - \mathbf{s_j})dt + o(dt)\right) + o(dt), \tag{3.6}$$

where for convenience $a_j(\mathbf{X}) = a_j(\mathbf{X}(t))$. By rewriting the previous statement to

$$p(\mathbf{X}, t + dt) - p(\mathbf{X}, t) = -p(\mathbf{X}, t)\sum_{j=1}^{m} a_j(\mathbf{X})dt + \sum_{j=1}^{m} p(\mathbf{X} - \mathbf{s_j}, t)a_j(\mathbf{X} - \mathbf{s_j})dt + o(dt), \tag{3.7}$$

by dividing both sides by dt, and letting $dt \to 0$, one receives the general chemical master equation.

Theorem 3.4. *The chemical master equation to a chemical reaction network as stated in equation (3.1) is given by*

$$\frac{d}{dt}p(\mathbf{X}, t) = -p(\mathbf{X}, t)\sum_{j=1}^{m} a_j(\mathbf{X}) + \sum_{j=1}^{m} p(\mathbf{X} - \mathbf{s_j}, t)a_j(\mathbf{X} - \mathbf{s_j}). \tag{3.8}$$

Adapted from [72].

The chemical master equation comprises a system of coupled linear differential equations; one linear differential equation for every possible state $\mathbf{X}(t)$ of the system. Hence, when the state space of a reaction network is large or infinite it is infeasible to evaluate the exact solution of the corresponding chemical master equation [73]. Only for very simple chemical reaction networks [74], when steady-state distributions are assumed [32] or in restricted regimes [75], may an analytical solution of the chemical master equation be determined. Systems of ordinary or stochastic differential equations present one option of approximating the solution of the chemical master equation. However, the chemical master equation only reduces to these approximations in the thermodynamic limit i.e., where the number of molecules of all chemical species and the containing volume approach infinity such that the molecular concentrations continue to be finite values [76]. Other approximation methods contain but are not limited to the Fokker-Planck approximation [77], the T-factor method [78], finite state projection [79] or van Kampens linear noise approximation [80]. However, if the exact solution to the chemical reaction network is of importance and its intrinsic stochastic nature should be maintained, one reverts to stochastic simulation algorithms.

3.4. Gillespie's stochastic simulation algorithm

In 1976, Daniel T. Gillespie proposed a method to numerically simulate the stochastic time evolution of a chemical reaction network [81], which he refined in 1977 [82]. The general idea is to repeatedly sample from the stochastic process $\mathcal{X}$ to approximate the solution to the chemical master equation. Therefore, Gillespie defined a reaction probability density function [81]

$$\begin{aligned} p(\tau, j|\mathbf{X}, t)d\tau = {} & \text{probability that the next reaction to occur is reaction } R_j \text{ and} \\ & \text{happens in the infinitesimal interval } [t + \tau, t + \tau + d\tau) \text{ given that} \\ & \text{the current system is in state } \mathbf{X}(t). \end{aligned}$$

In comparison to the reaction propensity defined in Definition 3.5, the reaction probability function also contains information about the waiting time. Similar to the derivation of the

chemical master equation, this is the product of the probability $p_1(\tau|\mathbf{X},t)$ that no reaction occurs in time interval $[t, t+\tau]$ and the probability $p_2(\tau|\mathbf{X},t)d\tau$ that reaction R_j occurs in time interval $[t+\tau, t+\tau+d\tau]$. This is given by

$$\begin{aligned} p(\tau, j|\mathbf{X},t)d\tau =& p_1(\tau|\mathbf{X},t)p_2(\tau|\mathbf{X},t)d\tau \\ =& e^{-\sum_{j=1}^{m} a_j(\mathbf{X}(t))\tau} a_j\mathbf{X}(\mathbf{t}), \end{aligned}$$

with $e^{-\sum_{j=1}^{m} a_j(\mathbf{X}(t))\tau}$ being the evaluation of the Poisson distribution density with rate $a_j(\mathbf{X}(t))\tau$ at 0 and $a_j\mathbf{X}(\mathbf{t})$ the reaction propensity of reaction R_j. When defining

$$a_0(\mathbf{X}(t)) = \sum_{j=1}^{m} a_j(\mathbf{X}(t)) \tag{3.9}$$

it becomes apparent that $p(\tau, j|\mathbf{X},t)d\tau$ is a joint density of two independent random variables

$$p(\tau, j|\mathbf{X},t)d\tau = a_0(\mathbf{X}(t))e^{-a_0(\mathbf{X}(t))\tau}\frac{a_j(\mathbf{X}(\mathbf{t}))}{a_0(\mathbf{X}(\mathbf{t}))}, \tag{3.10}$$

more specifically, an exponential random variable with rate parameter $a_0(\mathbf{X}(i))$ which can be interpreted as the time distribution to the next reaction

$$T(\mathbf{X}(t)) = a_0(\mathbf{X}(t))e^{-a_0(\mathbf{X}(t))}, \tag{3.11}$$

and a discrete random variable representing the discrete index distribution I given by

$$\mathbb{P}(I(\mathbf{X}(t) = j)) = \frac{a_j(\mathbf{X}(t))}{a_0(\mathbf{X}(t))} \qquad \forall j = 1 : m. \tag{3.12}$$

Hence, one can simply draw samples from the reaction probability density function. Today this method is known as Gillespie's stochastic simulation algorithm or Gillespie's next reaction method, which correctly accounts for the intrinsic stochasticity of a reaction system and its correlations. The pseudo-code for the approach can be found in Algorithm 1.

Algorithm 1 Gillespie's stochastic simulation algorithm

Input: initial state $\mathbf{X}_0 \in \Gamma$, reaction propensities $a_j(\mathbf{x}(t))$ for each $R_j \in \mathcal{R}$ and stoichiometric matrix S

Output: trajectory of the stochastic process $\mathcal{X}$

```
set t(0) ← 0
for i = 0, 1, 2, ... do
    sample τ ~ T(X(i))
    sample j ~ I(X(i))
    t(i+1) ← t(i) + τ
    X(i+1) ← X(i) + s_j
    record (X(i+1), t(i+1))
end for
```

Example. Two realizations of the leaky telegraph model with identical initial states using Gillespie's stochastic simulation algorithm are shown in Figure 3.2.

However, with a large number of chemical species $\mathcal{C}$ or reactions $\mathcal{R}$ the cumulative propensity $a_0(\mathbf{X}(i))$ increases, rendering the computation of the exact solution to the chemical master equation infeasible [83]. In order to overcome these limitations, approximation methods for the simulation of stochastic systems have been introduced. A well known approach in the field is the τ-leaping method, where the propensity function is updated only after fixed time intervals [83].

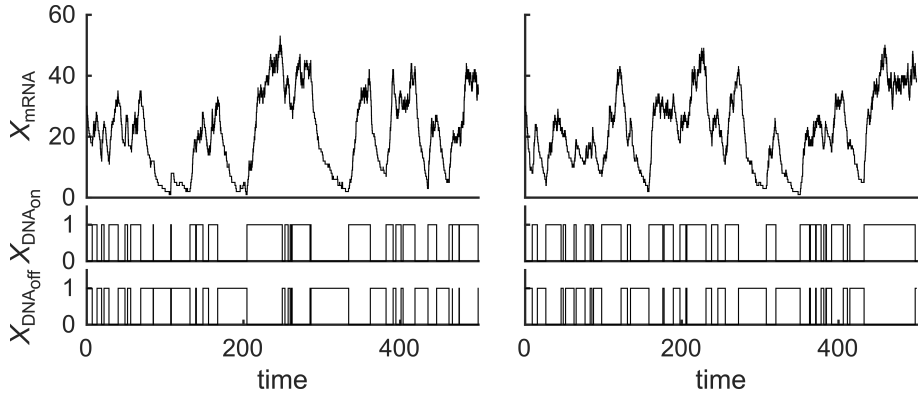


Figure 3.2: Two exemplary realizations of the leaky telegraph model with identical initial states. States [$X_{\mathrm{DNA_{off}}}$ (bottom), $X_{\mathrm{DNA_{on}}}$ (middle), X_{mRNA} (top)] are shown for each realization with identical initialization $\mathbf{X}_0 = [1, 0, 30]$ and rate constants $r_{\mathrm{on}} = 0.1$, $r_{\mathrm{off}} = 0.1$, $r_{\mathrm{prod}} = 0.2$, $d = 20$ and $r_{\mathrm{deg}} = 0.1$.

Other approximation methods include, but are not limited to, the Langevin method [84], the Poisson approximation method [85], multi-scale stochastic simulations [86, 87] and the integration of time-dependent rate constants [88].

3.5. Application: Intrinsic noise and rare, transient gene expression regulation

We give a brief example of a stochastic gene expression model, which integrates previously introduced concepts from graph theory (section 2) and stochastic modeling. We would like to emphasize that this section focuses on the brief description of the development of a stochastic gene expression model used to study the effect of intrinsic noise on transient and rare gene expression regulation. The details and the full study can be found in A.1. This section is based on (and is partly identical to) core publication [1].

Research problem Melanoma cells have demonstrate rare transient coordinated high gene expression. By 'coordinated high expression' we mean the simultaneous high expression in a subset of genes for a single cell and where by 'transient' we mean the appearance and disappearance of coordinated high expression within a cell over time (Figure 3.3) [47]. Only a rare subpopulation of cells shows coordinated high gene expression at a given time leading to heavy-tailed expression distributions in these genes at the population level. Moreover, only cells with coordinated high expression at the time of drug administration are thought to be pre-resistant, surviving drug therapy and acquiring stable resistance. It is thus of clinical relevance to identify the origins of this rare transient coordinated high expression state. In other words, is intrinsic noise sufficient to produce these rare transient coordinated high expression states?

Approach To investigate the emergence of rare transient coordinated high expression states, we developed a mathematical framework to simulate and evaluate gene expression patterns. The mathematical framework consists of a set of gene regulatory networks, a minimal mathematical model that recapitulates the observed gene expression patterns and stochastic simulations.

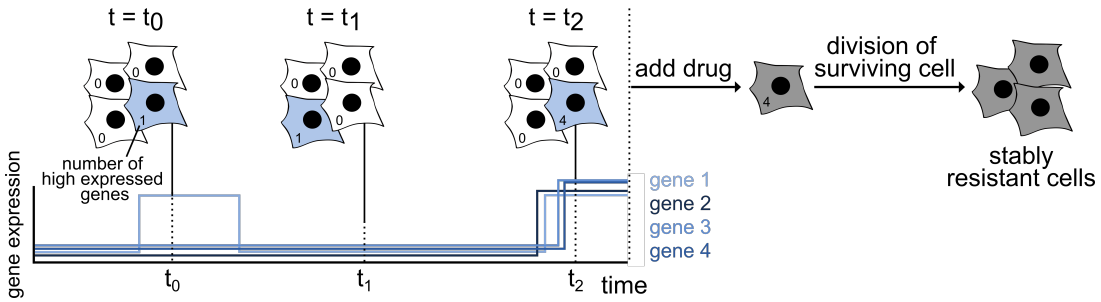


Figure 3.3: Drug-naive melanoma cells exist in low (white cells) as well as rare transient coordinated high (blue cells) expression states. Cells in the rare coordinated high state are pre-resistant. A schematic of the corresponding expression pattern is shown in the lower panel. The cells in a high expression state are more likely to survive and acquire resistance upon drug administration. Adapted from [1].

Gene regulatory networks Experimental data of Shaffer et al. [47] shows that the frequency of cells with coordinated high expression exceeds the expected frequency when assuming the events of high expression per gene to be mutually independent. To explain the observed coordination between the high expressions of several genes, we assumed gene regulatory effects to be an integral part of our model. Using graph theoretical properties, we defined a set of suitable gene regulatory networks for which we performed simulations and our downstream analysis. To efficiently and systematically evaluate gene regulatory networks of different sizes in an unbiased way, we restricted our analysis to the subset of weakly connected, non-isomorphic and symmetric digraphs (see Chapter 2). This reduced the testable space of unique gene regulatory networks by several orders of magnitude. We further restricted the analysis to digraphs of orders $2, 3, 5$ and 8, comprising $2, 4, 10$ and 80 weakly connected, non-isomorphic and symmetric digraphs, respectively. In sum, we analyzed 96 different gene regulatory networks.

Transcriptional bursting model We used an extended version of the leaky telegraph model (see previous examples) to describe gene expression of multiple genes within a pre-resistant melanoma cell. The gene regulations are described according to predefined gene regulatory networks (see previous paragraph). We assumed that gene regulation modulates the reaction propensity of DNA activation of the regulated gene. The strength of the regulation is modeled by a Hill function (section 4.3) and depends on the current mRNA count of the regulating gene. Correspondingly, the reaction propensity of the DNA activation, so far denoted by $a_1(\mathbf{X}(t))$, of a regulated gene G_j is modified from

$$a_{1Gj}(\mathbf{X}(t)) = r_{\text{on}Gj} X_{\text{DNAoff}G_j}(t) \tag{3.13}$$

to

$$a_{1Gj}(\mathbf{X}(t)) = \left(r_{\text{on}Gj} + \sum_{i=1}^{|V|} r_{\text{add}\,GiGj} \frac{X_{\text{mRNA}Gi}^{n_{GiGj}}(t)}{k_{GiGj}^{n_{GiGj}} + X_{\text{mRNA}Gi}^{n_{GiGj}}(t)} \right) X_{\text{DNAoff}G_j}(t), \tag{3.14}$$

where $r_{\text{add}\,GiGj} = 0 \; \forall G_i \in \mathcal{V} : (G_i, G_j) \notin \mathcal{E}$ describes the regulatory strength of gene G_i on the activation of gene G_j and $k_{GiGj}, n_{GiGj} > 0 \; \forall G_i, G_j \in \mathcal{V}$ parameters of the Hill function (see equation (4.4)).

Example. For the gene regulation depicted in Figure 3.4A the reaction propensity of the DNA

activation reaction, so far denoted by $a_1(\mathbf{X}(t))$, of the regulated gene G_2 is given by

$$a_{1G2}(\mathbf{X}(t)) = \left(r_{\text{on}G2} + r_{\text{add}\,G1G2} \frac{X_{\text{mRNA}G1}^{n_{G1G2}}(t)}{k_{G1G2}^{n_{G1G2}} + X_{\text{mRNA}G1}^{n_{G1G2}}(t)} \right) X_{\text{DNAoff}G2}(t). \tag{3.15}$$

The corresponding transcriptional bursting model for genes G_1 and G_2 is shown in Figure 3.4B.

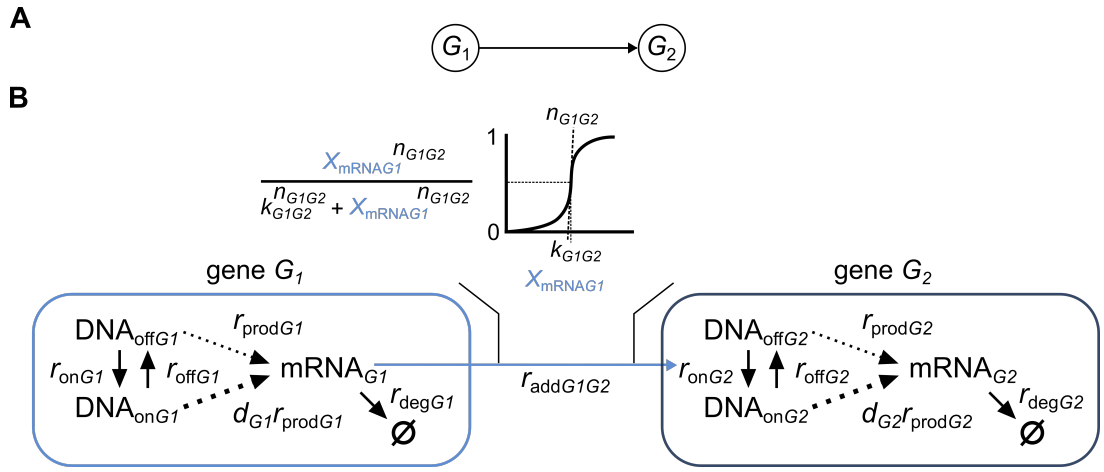


Figure 3.4: Transcriptional bursting model. (A) Exemplary gene regulatory network defining the gene regulation between genes G_1 and G_2, where gene G_1 is the regulating gene and gene G_2 the regulated gene. (B) Schematic of the transcriptional bursting model for two genes G_1 and G_2. Gene regulation is modeled by a Hill function, where the gene expression of the regulating gene G_1 increases the activation of DNA of regulated gene G_2. The dotted arrows represent chemical reactions without conversions. Adapted from [1].

The Hill function within the modified reaction propensity (equation (3.14)) summarizes a series of microscopic reactions of which each in itself complies with the law of mass action 3.2. The Hill function is used to reduce the complexity of the system and to allow for an efficient evaluation of the model [89]. Although, the model itself now no longer complies with the law of mass action, the chemical master equation remains valid as long as the modified reaction propensity is interpreted as the firing rate of the corresponding reaction [90]. We further assumed that translation is fast such that mRNA is modeled to directly regulate gene expression, thereby allowing us to neglect the protein level and to reduce model complexity. Furthermore, we restricted our model to additive and activating gene regulations only.

Simulation parameters To reduce the number of model parameters and to allow for the comparison of simulated expression patterns between gene regulatory networks, we assumed all genes to be regulated by the same parameter vector Θ given by $\Theta = \{r_{\text{on}}, r_{\text{off}}, r_{\text{prod}}, d, r_{\text{deg}}, r_{\text{add}}, n\}$, with $r_{\text{add}} = r_{\text{add}GiGj}$ and $n = n_{GiGj}$ $\forall G_i, G_j \in V$ and the dissociation constant k is defined as

$$k = 0.95 \frac{dr_{\text{prod}}}{r_{\text{deg}}}, \tag{3.16}$$

where $\frac{dr_{\text{prod}}}{r_{\text{deg}}}$ the steady-state solution of the deterministic description of this chemical reaction network assuming $X_{\text{DNA}_{\text{on}}}(t) = 1$ $\forall t$. We defined the dissociation constant k as above to restrict the emergence of coordinated high expression. 1,000 parameter sets, $\Theta_1, ..., \Theta_{1,000}$, were then sampled according to the latin-hypercube sampling method [91]. For more details on the simulation parameters and parameter boundaries see [1].

Stochastic simulations Using the chemical reaction network described by the transcriptional bursting model and Gillespie's stochastic simulation algorithm (see section 3.4), we simulated the gene expression for all 96 gene regulatory networks and $1,000$ parameter sets, resulting in $96,000$ simulations, for $1,000,000$ time units. An exemplary frame of a simulation is shown in Figure 3.5.

Quantitative criteria To identify simulations with rare coordinated high states, we defined a set of quantitative criteria describing the rare coordinated high expression patterns as observed in pre-resistant melanoma cells. We first determined a threshold of expressed mRNA counts above which we call a gene highly expressed.

Definition 3.6. *A gene $G_i \in \mathcal{V}$ is called highly expressed, if*

$$X_{\mathrm{mRNA}Gi}(t) > 0.8\frac{dr_{\mathrm{prod}}}{r_{\mathrm{deg}}}, \tag{3.17}$$

where d, r_{prod}, and r_{deg} are corresponding simulation parameters.

The threshold is set to 80% of the maximal mean expression, assuming a constantly active DNA state given by $\frac{dr_{\mathrm{prod}}}{r_{\mathrm{deg}}}$. Next, we wanted the simulated gene expressions of multiple genes to show coordinated high expression, i.e., more than half of the genes of a given gene regulatory network express mRNA counts above the previously defined threshold for high expression, at least once throughout the simulation. Furthermore, we wanted coordinated high expression to be a rare phenomenon across a pre-resistant melanoma cell population. Hence, we created an artificial cell population of $1,000$ cells of $1,000$ time units from the original $1,000,000$ time units simulation. For the artificially created cell population, we then determined the mRNA counts of all genes for time point t_{rand} to evaluate whether the expression distributions at the population level show characteristics of heavy-tailed distributions as identified by Shaffer et al. [47].

Definition 3.7. *A simulation shows rare coordinated high expression if*

(i) (coordinated high expression) $\exists \mathcal{W} \subseteq \mathcal{V} : |\mathcal{W}| > \frac{|\mathcal{V}|}{2}$ and $\exists t_k : X_{\mathrm{mRNA}Gi}(t_k)$ highly expressed $\forall G_i \in \mathcal{W}$ and

(ii) (rare) for a randomly determined $t_{\mathrm{rand}} \in 1, ..., 999$ the distributions $\{X_{\mathrm{mRNA}Gi}(q)\}_{\{q=Q\}}$, with $Q = \{t_{\mathrm{rand}} : 1,000 : 999,000 + t_{\mathrm{rand}}\}$ $\forall G_i \in \mathcal{V}$, are right-skewed and unimodal.

An exemplary frame of a simulation showing rare coordinated high states and a comparison between an experimental and a simulated expression distribution are shown in Figure 3.5.

Analysis and results Using the mathematical framework and the resulting simulations as described above, we

(i) found that the transcriptional bursting model based on established principles of transcription and gene regulation is sufficient to describe the origin of rare coordinated high expression.

(ii) observed that $< 1\%$ of the $96,000$ simulations show rare coordinated high gene expression.

(iii) showed that all tested gene regulatory network sizes were able to produce simulations with rare coordinated high expression.

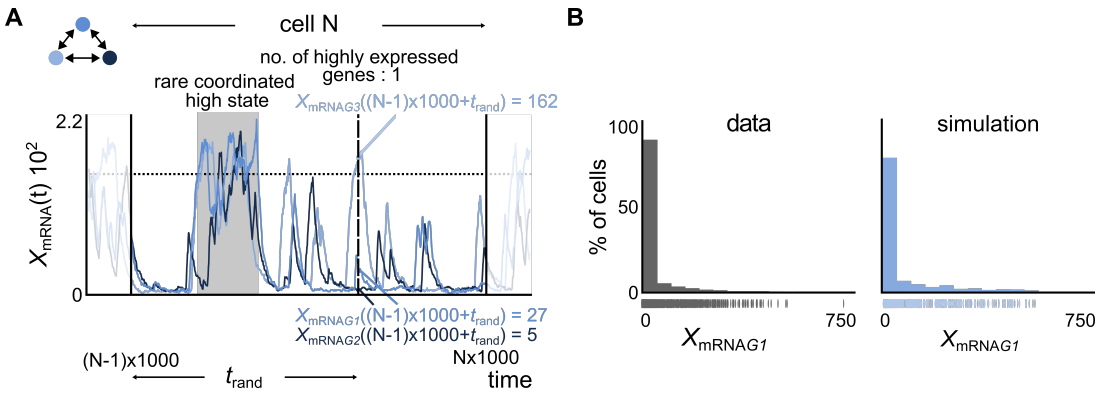


Figure 3.5: Example simulation of the transcriptional bursting model. (A) Exemplary frame of a simulation with a rare coordinated high state (highlighted in gray) for a network of order 3 (top left). The expression of an artificially created cell N is shown and the criteria evaluated for rare coordinated high expression are highlighted. The dotted line represents the threshold for high expression. (B) The simulated expression distribution at the population level is qualitatively similar to the experimental data from a drug-naive melanoma population. Adapted from [1].

(iv) demonstrated that within a given network size the number of simulations with rare coordinated high expression decreases with increasing in-degree, where by definition all vertices in a symmetric gene regulatory network have the same in-degree.

(v) identified a set of parameters which is particularly favorable in producing simulations with rare coordinated high expression ($> 20\%$ of the simulations).

(vi) performed a decision tree analysis and found that only 3 out of the 7 model parameters, $r_{\text{on}}, r_{\text{off}}$ and r_{add}, show a strong correlation with the parameter sets particularly favorable in producing simulations with rare coordinated high expression.

(vii) found a constrained parameter space based on the decision tree analysis for which the frequency of simulations with rare coordinated high expression is increased $>$ 10-fold as compared to the original parameter space.

(viii) studied the initiation, maintenance and termination of rare coordinated high expression states and found that a long transcriptional burst initiates, an increase in burst frequency maintains, and a random process terminates rare coordinated high expression states.

(ix) hypothesized that stable resistance after drug administration may be achieved by an increase in the number of edges present in the gene regulatory network which was validated experimentally by comparing the number of edges of inferred gene regulatory networks for drug-naive cells and resistant melanoma colonies.

Conclusion In sum, we established a minimal gene expression model comprising transcriptional bursting, gene regulation and stochasticity, which recapitulates the gene expression patterns observed in pre-resistant melanoma cells. More generally speaking, we showed that intrinsic noise and established principles of gene expression regulation are sufficient to lead to rare and transient gene expression patterns.

4. Deterministic gene expression models

In the previous chapter, we showed that stochasticity in gene expression is a driver of cell-to-cell variability. However, when considering e.g., the mean expression of a large cell population, stochastic effects no longer dominate the expression dynamics. In this case, a deterministic description is more simple and appropriate.

Initially, the stochastic and deterministic descriptions of chemical reaction networks were studied independently. Work by Thomas G. Kurtz in the 1970s demonstrated that the stochastic model converges to the deterministic description in the thermodynamic limit [76, 92].

In Schuh et al. [2], we developed a deterministic model describing the histone modification kinetics in frog embryos during early development. We used this model to identify whether external cues can be an integral part of gene expression regulation. Moreover, in Schuh et al. [3], we considered a deterministic model of gene expression to describe the repression kinetics in yeast cells to study the effects of repeated external cues on the gene expression regulation of repression. This chapter is mainly based on [70, 93–95], if not stated otherwise.

4.1. Deterministic description of the chemical master equation

In Theorem 3.4, we introduced the chemical master equation, a potentially infinite system of ODEs describing the time evolution of a chemical reaction network. However, when only interested in the evolution of the mean, the system of ODEs reduces to a simpler, deterministic description.

The mean state of a chemical species X_i at time t can be described by

$$\mathbb{E}[X_i(t)] = \sum_{\mathbf{X}} X_i p(\mathbf{X}, t),$$

where $p(\mathbf{X}, t)$ the probability to be in state $\mathbf{X}$ at time t. Then the time evolution of the mean of X_i is given by

$$\frac{d}{dt}\mathbb{E}[X_i(t)] = \sum_{\mathbf{X}} X_i \frac{d}{dt} p(\mathbf{X}, t),$$

where we make use of the previously derived chemical master equation (see Definition 3.4):

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X_i(t)] &= \sum_{\mathbf{X}} X_i \left(-p(\mathbf{X}, t) \sum_{j=1}^{m} a_j(\mathbf{X}) + \sum_{j=1}^{m} p(\mathbf{X} - \mathbf{s_j}, t) a_j(\mathbf{X} - \mathbf{s_j}) \right) \\ &= -\sum_{\mathbf{X}} \sum_{j=1}^{m} X_i a_j(\mathbf{X}) p(\mathbf{X}, t) + \sum_{\mathbf{X}} \sum_{j=1}^{m} X_i a_j(\mathbf{X} - \mathbf{s_j}) p(\mathbf{X} - \mathbf{s_j}, t).\end{aligned}$$

To further simplify the expression, we substitute $\mathbf{X}$ for $\mathbf{X} + \mathbf{s_j}$ in the right sum, resulting in

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X_i(t)] &= -\sum_{\mathbf{X}} \sum_{j=1}^{m} X_i a_j(\mathbf{X}) p(\mathbf{X}, t) + \sum_{\mathbf{X}} \sum_{j=1}^{m} (X_i + s_{ij}) a_j(\mathbf{X}) p(\mathbf{X}, t) \\ &= \sum_{j=1}^{m} s_{ij} \sum_{X} a_j(\mathbf{X}) p(\mathbf{X}, t) \\ &= \sum_{j=1}^{m} s_{ij} \mathbb{E}[a_j(\mathbf{X})].\end{aligned}$$

Hence, the evolution of the mean of a chemical reaction network is dependent on the stoichiometric matrix (see Definition 3.2) and the reaction propensities (see Definition 3.5), which are easily determined. The deterministic formulation of the evolution of the mean of a chemical reaction network is given by

$$\frac{d}{dt}\mathbb{E}[X_i(t)] = \sum_{j=1}^{m} s_{ij}\mathbb{E}[a_j(\mathbf{X})] \tag{4.1}$$

for chemical species C_i. However, this only holds for zero and first order reactions [89]. More specifically, for zero order reactions, i.e., reactions with constant reaction propensity $a_j(\mathbf{X}) = k$, $\mathbb{E}[a_j(\mathbf{X})] = k$ (see section 3.2), and for first order reactions, i.e., reactions with a linear reaction propensity $a_j(\mathbf{X}) = kX_i(t)$ (see section 3.2), $\mathbb{E}[a_j(\mathbf{X})] = k\mathbb{E}[X_i(t)]$. However, for second order reactions of different educts, i.e. with quadratic reaction propensity $a_j(\mathbf{X}) = kX_i(t)X_j(t)$ (see section 3.2), the expectation of $a_j(\mathbf{X})$ is given by

$$\mathbb{E}[a_j(\mathbf{X})] = a_j(\mathbb{E}[\mathbf{X}]) + \frac{1}{2}\sum_{l_1,l_2}\frac{\partial^2 a_j(\mathbf{X})}{\partial X_{l_1}\partial X_{l_2}}\mathbb{E}\left[(X_{l_1} - \mathbb{E}[X_{l_1}])(X_{l_2} - \mathbb{E}[X_{l_2}])\right], \tag{4.2}$$

having applied a Taylor series expansion at $\mathbb{E}[\mathbf{X}]$ and where $\mathbb{E}\left[(X_{l_1} - \mathbb{E}[X_{l_1}])(X_{l_2} - \mathbb{E}[X_{l_2}])\right]$ the second moment, which is itself dependent on higher order moments. Hence, if non-linear propensities are described, the equation is typically coupled to higher-order moments. If the system contains non-linear propensities, moment-closure approximations are used to decouple lower- and higher-order moments. Common methods are the third order moment-closure approximation, where the third and higher order moments are assumed to be zero [96], low dispersion closure, where the distribution around lower order moments is considered before assuming higher order moments to be negligible [97] or the mean-field closure [98]. Similar descriptions of the evolution of higher order moments of the chemical master equation can be derived [95].

Example. For the leaky telegraph model (see example of 3.1) the evolution of the mean for each of the chemical species $\mathcal{C} = \{\mathrm{DNA_{off}}, \mathrm{DNA_{on}}, \mathrm{mRNA}\}$ is described by the following ODEs:

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}[X_{\mathrm{DNA_{off}}}(t)] &= r_{\mathrm{off}}X_{\mathrm{DNA_{on}}}(t) - r_{\mathrm{on}}X_{\mathrm{DNA_{off}}}(t)\\
\frac{d}{dt}\mathbb{E}[X_{\mathrm{DNA_{on}}}(t)] &= r_{\mathrm{on}}X_{\mathrm{DNA_{off}}}(t) - r_{\mathrm{off}}X_{\mathrm{DNA_{on}}}(t)\\
\frac{d}{dt}\mathbb{E}[X_{\mathrm{mRNA}}(t)] &= r_{\mathrm{prod}}X_{\mathrm{DNA_{off}}}(t) + dr_{\mathrm{prod}}X_{\mathrm{DNA_{on}}}(t) - r_{\mathrm{deg}}X_{\mathrm{mRNA}}(t).
\end{aligned}$$

4.2. Michaelis-Menten enzyme kinetics

A special case of chemical reactions are enzyme reactions. Enzymes catalyze biochemical reactions but are not consumed or altered by the process. In 1913, Leonor Michaelis and Maud Menten proposed a mathematical model to describe the rate of a general enzymatic reaction and its product formation [99]. This section was adapted from [99, 100].

Theorem 4.1. *Let $\mathcal{C} = \{E, S, ES, P\}$ be chemical species, where E is the enzyme, S the substrate, ES the enzyme-substrate complex, and P the product, and let the chemical reaction network be described by*

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E} + \mathrm{P}.$$

Then the rate at which the product concentration $[P]$ is formed is given by

$$\frac{d}{dt}[P] = \upsilon = \upsilon_{max}\frac{[S]}{k + [S]}, \tag{4.3}$$

where $[S]$ is the concentration of the substrate, k the dissociation constant specifying the substrate concentration at which the product formation is half-maximal, and υ_{max} is the maximal production rate. This relation is called the Michaelis-Menten kinetics.

In enzyme kinetics, the product formation rate can reach saturation. Even for high substrate concentrations the reaction rate is limited by the enzymatic reaction during which the substrate is converted into the product. For the full derivation see [99, 101]. In contrast to basing the deterministic model description on the law of mass action (see section 3.2), the Michaelis-Menten enzyme kinetics are commonly used to describe non-linear processes [102].

4.3. Hill function

In 1910, Archibald Vivian Hill introduced the so called Hill function describing the binding of oxygen to hemoglobin [103]. Under specific conditions the Hill function can be viewed as an extension of the Michaelis-Menten kinetics modeling the positive cooperative binding of several substrates, S_1, ..., S_n, to an enzyme with multiple subunits and active sites [102]. This section was adapted from [101, 103].

Theorem 4.2. *Let $\mathcal{C} = \{E, S, SE, SSE, P\}$ the chemical species, where E is the enzyme, S the substrate, SE the enzyme-substrate complex with one substrate bound, SSE the enzyme-substrate complex with two substrates bound, and P the product, and let the chemical reaction network be described by*

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{SE} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E} + \mathrm{P},$$
$$\mathrm{S} + \mathrm{SE} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{SSE} \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} \mathrm{SE} + \mathrm{P}.$$

Assuming positive cooperativity, i.e., upon the binding of one substrate the conformation of the enzyme changes, thereby facilitating the easier binding of a second substrate; then the rate at which the product concentration $[P]$ is formed can be approximated by

$$\frac{d}{dt}[P] = \upsilon = \upsilon_{max}\frac{[S]^n}{k^n + [S]^n}, \tag{4.4}$$

where $[S]$ the concentration of the substrate S, k the dissociation constant specifying the substrate concentration at which the product formation is half-maximal, υ_{max} the maximal rate, and n the Hill coefficient which describes the binding affinity or the number of substrates an enzyme can bind to simultaneously. This relation is called the Hill function.

For $n = 1$ the Hill function reduces to the Michaelis-Menten kinetics. Furthermore, for $n > 1$ the Hill function is sigmoidal and for $n \to \infty$, it approximates a step function, with $\upsilon = 0$ for $[S] < k$ and $\upsilon = \upsilon_{max}$ for $[S] > k$ (Figure 4.1). The Hill function is frequently used in mathematical modeling to describe non-linear responses of the regulatory functions in gene regulatory networks as introduced in chapter 2 [102].

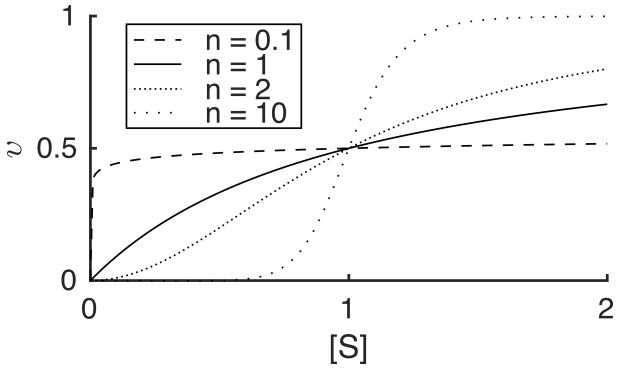


Figure 4.1: Exemplary Hill functions for $\upsilon_{max} = 1$, $k = 1$ and varying n.

Example. Let us consider two genes, G_1 and G_2, whose interactions are summarized by the gene regulatory network as shown in Figure 4.2 (see chapter 2).

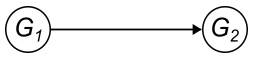


Figure 4.2: Gene regulatory network for genes G_1 and G_2, where G_1 activates G_2.

Let genes G_1 and G_2 result in proteins P_1 and P_2, respectively. Focusing on the protein level, let the chemical reaction network of chemical species $\mathcal{C} = \{\mathrm{P}_1,\mathrm{P}_2\}$ be described by reactions

$$\begin{aligned}
\mathrm{P}_1 \text{ production} \quad R_1\text{:} \quad & \emptyset \xrightarrow{r_{\mathrm{prod1}}} \mathrm{P}_1 \\
\mathrm{P}_1 \text{ degradation} \quad R_2\text{:} \quad & \mathrm{P}_1 \xrightarrow{r_{\mathrm{deg1}}} \emptyset \\
\mathrm{P}_2 \text{ production} \quad R_3\text{:} \quad & \emptyset \xrightarrow{r_{\mathrm{prod2}}} \mathrm{P}_2 \\
\mathrm{P}_2 \text{ degradation} \quad R_4\text{:} \quad & \mathrm{P}_2 \xrightarrow{r_{\mathrm{deg2}}} \emptyset.
\end{aligned}$$

Then, the state of the system is summarized by $\mathbf{X}(t) = [X_{\mathrm{P}_1}(t), X_{\mathrm{P}_2}(t)]$. Often, the reaction propensity corresponding to the production of the regulated gene P_2 and reaction R_3 is not described according to the law of mass action but accounts for a non-linear regulation of P_2 by P_1:

$$a_3(\mathbf{X}(t)) = r_{\mathrm{prod2}} \frac{X_{\mathrm{P}_1}(t)^{n_{12}}}{k_{12}^{n_{12}} + X_{\mathrm{P}_1}(t)^{n_{12}}},$$

where k_{12} is the dissociation constant and n_{12} the Hill coefficient of the corresponding Hill function describing the non-linearity. The Hill function summarizes a series of microscopic reactions of which each in itself complies with the law of mass action (see 3.2) [89]. Although the model itself now no longer complies with the law of mass action, the chemical master equation remains valid as long as the modified reaction propensity is interpreted as the firing rate of the corresponding reaction [90].

4.4. Deterministic model formulation

So far, we focused on the deterministic formulation describing a biological process. However, we are often not able to measure the states of a biological system directly. Hence, when using the deterministic description to explain a biological process, we need to couple the true underlying dynamical process to a function describing the measurable states.

Definition 4.1. *A function of the states and/or parameters, $\boldsymbol{\theta}$, which can be measured, are*

called observables,

$$\mathbf{y}(t, \boldsymbol{\theta}) = h\left(t, \boldsymbol{\theta}, \mathbf{x}(t, \boldsymbol{\theta})\right), \tag{4.5}$$

with observation function $h : \mathbb{R} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^n \to \mathbb{R}^{n_y}$ *and* n_y *the number of observables.*

Example. For the leaky telegraph model (see example of 3.1), we have previously described the deterministic formulation of the dynamic process. However, we cannot measure all of the states, $\mathbf{X}(t) = [X_{\mathrm{DNA_{off}}}(t), X_{\mathrm{DNA_{on}}}(t), X_{\mathrm{mRNA}}(t)]$ for time point t. mRNA concentrations of a pre-defined gene can be detected in single, fixed cells, by a technique referred to as RNA fluorescence in situ hybridization (RNA FISH) [104]. Here, fluorescent probes are hybridized to mRNA molecules. The signal is then detected under the microscope. Hence, in this example, only the mRNA concentrations, $X_{\mathrm{mRNA}}(t)$ for time point t, can be experimentally measured and the observable is described by

$$y(t) = X_{\mathrm{mRNA}}(t).$$

A deterministic model couples the deterministic description of a dynamic process and the observables.

Definition 4.2. *A deterministic model* $\mathcal{M}$ *is given by a deterministic description of the system dynamics* $\frac{\partial}{\partial t}\mathbf{x}(t, \boldsymbol{\theta})$ *and observables* $\mathbf{y}(t, \boldsymbol{\theta})$ *such that*

$$\mathcal{M}(\boldsymbol{\theta}) = \begin{cases} \frac{\partial}{\partial t}\mathbf{x}(t, \boldsymbol{\theta}) = f\left(t, \boldsymbol{\theta}, \mathbf{x}(t, \boldsymbol{\theta})\right), \\ \mathbf{y}(t, \boldsymbol{\theta}) = h\left(t, \boldsymbol{\theta}, \mathbf{x}(t, \boldsymbol{\theta})\right), \end{cases} \tag{4.6}$$

with initial states $\mathbf{x}(t_0, \boldsymbol{\theta}) = \mathbf{x}_0(\boldsymbol{\theta})$*, where* $\boldsymbol{\theta}$ *is the parameter vector,* $f : \mathbb{R} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^n \to \mathbb{R}^n$*, and* $h : \mathbb{R} \times \mathbb{R}^{n_\theta} \times \mathbb{R}^n \to \mathbb{R}^{n_y}$ *the observation function.*

As the system $\mathbf{x}(t, \boldsymbol{\theta})$ is a function of time t and parameters $\boldsymbol{\theta}$, we here use the partial derivative $\frac{\partial}{\partial t}$.

5. Parameter estimation

For most biological processes, the parameter vector $\boldsymbol{\theta}$ of the underlying model $\mathcal{M}(\boldsymbol{\theta})$ (see Definition 4.2) is unknown. However, to understand aspects of a biological process and to use the model for predictions, one is interested in identifying the underlying parameters. As $\boldsymbol{\theta}$ can often not be directly measured, the parameters need to be inferred from experimental data. In both publications [1, 2] and the manuscript [3], we aimed at identifying the model parameters underlying the varying biological processes. Overall, there are several approaches to parameter estimation including the frequentist [91] and Bayes approaches [105]. In this thesis, we mainly consider the former. This chapter is mainly based on [70, 106–108], if not stated otherwise.

5.1. Mathematical formulation

The general idea of parameter estimation is to find an estimated parameter vector $\boldsymbol{\theta}^*$ for which the forward simulation $\mathcal{M}(\boldsymbol{\theta}^*)$ describing the system states $\mathbf{x}(t, \boldsymbol{\theta}^*)$ and observations $\mathbf{y}(t, \boldsymbol{\theta}^*)$ best captures the experimental data $\mathcal{D}$. This can be formulated as a minimization problem, where the optimal parameter vector under $\mathcal{M}(\boldsymbol{\theta})$, $\boldsymbol{\theta}^*$, is given by

$$\boldsymbol{\theta}^* = \underset{\substack{\boldsymbol{\theta}\in\Theta \\ \text{subject to } \mathcal{M}(\boldsymbol{\theta})}}{\arg\min} \; J(\mathbf{y}(t, \boldsymbol{\theta}), \mathbf{y}^{\mathcal{D}}(t)), \tag{5.1}$$

where $J(\boldsymbol{\theta})$ the objective function allowing for a quantitative comparison between simulation $\mathbf{y}(t, \boldsymbol{\theta})$ and observations $\mathbf{y}^{\mathcal{D}}(t)$, and Θ a suitable parameter space. The optimal parameter vector $\boldsymbol{\theta}^*$ is hence dependent on the choice of the objective function $J(\mathbf{y}(t, \boldsymbol{\theta}), \mathbf{y}^{\mathcal{D}}(t))$.

5.2. Naive approach

In Schuh et al. [2], we were interested in developing a stochastic gene expression model that is able to recapitulate the gene expression patterns observed in pre-resistant melanoma cells. Due to the stochastic nature of our model, we applied a naive approach to parameter estimation as outlined by Algorithm 2.

Algorithm 2 Naive parameter estimation approach

Input: model $\mathcal{M}(\boldsymbol{\theta})$, parameter space Θ, the number of to-be sampled parameter vectors N, and data observations $\mathbf{y}^{\mathcal{D}}(t)$
Output: estimated optimal parameter vector $\boldsymbol{\theta}^*$
  for $i = 1, ..., N$ **do**
    sample $\boldsymbol{\theta}_i \in \Theta$ by e.g., latin-hypercube sampling
    forward simulate $\mathcal{M}(\boldsymbol{\theta}_i)$ to get $\mathbf{y}(t, \boldsymbol{\theta}_i)$
    compare $\mathbf{y}(t, \boldsymbol{\theta}_i)$ with $\mathbf{y}^{\mathcal{D}}(t)$ by pre-determined criteria
  end for
  $\boldsymbol{\theta}^* = \boldsymbol{\theta}_i$ for which $\mathbf{y}(t, \boldsymbol{\theta}_i)$ best describes $\mathbf{y}^{\mathcal{D}}(t)$

The aforementioned pre-determined criteria for comparing $\mathbf{y}(t, \boldsymbol{\theta}_i)$ with $\mathbf{y}^{\mathcal{D}}(t)$ do not necessarily need to be distance metrics. In Schuh et al. [1], we used several criteria to determine whether our biological phenomenon of interest was recapitulated by the stochastic simulations, e.g., whether any gene was highly expressed, the number of simultaneously highly expressed genes, and the gene expression distributions.

The performance of the naive approach is highly dependent on the number of parameter vectors sampled from the parameter space Θ. With increasing dimensionality of Θ, the coverage of the parameter space decreases exponentially, a phenomenon termed the curse of dimensionality. Hence, if possible, more advanced and efficient methods for estimating the parameter vector $\boldsymbol{\theta}^*$ should be used.

5.3. Likelihood-based approach: maximum likelihood estimation

One of the most common methods for parameter estimation is the likelihood-based approach, where the objective function of the minimization problem as defined in equation (5.1) is described by the (negative) likelihood. In Schuh et al. [2] and [3] we applied maximum likelihood estimation to infer the model parameters of our deterministic model descriptions. Generally, the likelihood function describes the probability of observing data set $\mathcal{D}$ given model $\mathcal{M}(\boldsymbol{\theta})$ and model parameters $\boldsymbol{\theta}$.

Definition 5.1. *The likelihood is defined as*

$$\mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}) = \mathbb{P}(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{n_t} \mathbb{P}(\mathbf{y}^{\mathcal{D}}(t_k)|\mathbf{y}(t_k,\boldsymbol{\theta}),\boldsymbol{\sigma}_k) \tag{5.2}$$

$$\overset{\text{independence of } y_i}{=} \prod_{k=1}^{n_t}\prod_{i=1}^{n_y} \mathbb{P}(y_i^{\mathcal{D}}(t_k)|y_i(t_k,\boldsymbol{\theta}),\sigma_{i,k}) \tag{5.3}$$

where $\mathcal{D} = \{((y_i^{\mathcal{D}}(t_k))_{i=1}^{n_y}, t_k)\}_{k=1}^{n_t}$ *the data set,* n_y *the number of observables* $\mathbf{y}(t_k,\boldsymbol{\theta})$*,* n_t *the number of time points,* $\mathbf{y}^{\mathcal{D}}(t_k)$ *the data measurements at time* t_k*,* $\boldsymbol{\theta}$ *a parameter vector describing* $\mathcal{M}(\boldsymbol{\theta})$ *and* $\boldsymbol{\sigma}_k$ *the noise parameters of an underlying noise model.*

In contrast to the previous chapter, where one was interested in minimizing the distance between the model simulation and the observations, we here want to identify the parameter vector having the highest probability of having produced data $\mathcal{D}$ under model $\mathcal{M}(\boldsymbol{\theta})$. Hence, we are interested in the parameter vector maximizing the likelihood function instead of minimizing the objective function.

Definition 5.2. *The maximum likelihood estimate* $\boldsymbol{\theta}^{ML}$ *is a solution to the optimization problem*

$$\boldsymbol{\theta}^{ML} = \underset{\substack{\boldsymbol{\theta}\in\Theta \\ \text{subject to } \mathcal{M}(\boldsymbol{\theta})}}{\arg\max} \mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}). \tag{5.4}$$

The maximum likelihood estimate is widely used due to its properties in the limit of large sample sizes, i.e., efficiency, consistency, and asymptotic normality. To improve the numerical solvability of the optimization problem one often rephrases the maximum likelihood estimate of equation (5.4) to

$$\boldsymbol{\theta}^{ML} = \underset{\substack{\boldsymbol{\theta}\in\Theta \\ \text{subject to } \mathcal{M}(\boldsymbol{\theta})}}{\arg\min} -\ell_{\mathcal{D}}(\boldsymbol{\theta}), \tag{5.5}$$

where

$$-\ell_{\mathcal{D}}(\boldsymbol{\theta}) = -\log \mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}) \tag{5.6}$$

$$= -\sum_{k=1}^{n_t}\sum_{i=1}^{n_y} \log \mathbb{P}(y_i^{\mathcal{D}}(t_k)|y_i(t_k,\boldsymbol{\theta}),\sigma_{i,k}) \tag{5.7}$$

is the negative log-likelihood [107]. When comparing the minimization problems of equations (5.1) and (5.5), it becomes obvious that the negative log-likelihood equals the objective function in the likelihood-based approach. If possible, log-transforming the model parameters $\boldsymbol{\theta}$ further improves the computational efficiency of the optimization [91, 109–111].

5.4. Noise models

As experimental data is generally noise corrupted, observed data points $y_i^{\mathcal{D}}(t_k)$ underlie a noise model, which directly defines the likelihood function (Definition 5.1).

Gaussian noise The most common noise model follows an additive Gaussian distribution, assuming

$$y_i^{\mathcal{D}}(t_k) = y_i(t_k, \boldsymbol{\theta}) + \epsilon_{i,k}, \qquad \epsilon_{i,k} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{i,k}^2(\boldsymbol{\theta})), \tag{5.8}$$

where the noise is assumed to be independently and identically distributed (iid). The negative log-likelihood is then given by

$$-\ell_{\mathcal{D}}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k=1}^{n_t} \sum_{i=1}^{n_y} \log(2\pi\sigma_{i,k}^2(\boldsymbol{\theta})) + \left(\frac{y_i^{\mathcal{D}}(t_k) - y_i(t_k, \boldsymbol{\theta})}{\sigma_{i,k}(\boldsymbol{\theta})} \right)^2. \tag{5.9}$$

Alternatively, one can assume the data to underlie a multiplicative log-Gaussian noise model where

$$y_i^{\mathcal{D}}(t_k) = y_i(t_k, \boldsymbol{\theta})\epsilon_{i,k}, \qquad \epsilon_{i,k} \overset{iid}{\sim} \log\mathcal{N}(0, \sigma_{i,k}^2(\boldsymbol{\theta})). \tag{5.10}$$

However, the multiplicative log-Gaussian noise model can be described by an additive Gaussian noise model when considering the log of the data points $y_i^{\mathcal{D}}(t_k)$, assuming all $y_i^{\mathcal{D}}(t_k) > 0$.

Laplacian noise When modeling outlier-corrupted data, one may rather use a Laplacian noise model accounting for heavier tails [112], with

$$y_i^{\mathcal{D}}(t_k) = y_i(t_k, \boldsymbol{\theta}) + \epsilon_{i,k}, \qquad \epsilon_{i,k} \overset{iid}{\sim} \text{Laplace}(0, \sigma_{i,k}^2(\boldsymbol{\theta})). \tag{5.11}$$

The negative log-likelihood of the additive Laplacian noise model is given by

$$-\ell_{\mathcal{D}}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k=1}^{n_t} \sum_{i=1}^{n_y} \log(2\sigma_{i,k}(\boldsymbol{\theta})) + \frac{|y_i^{\mathcal{D}}(t_k) - y_i(t_k, \boldsymbol{\theta})|}{\sigma_{i,k}(\boldsymbol{\theta})}. \tag{5.12}$$

5.5. Optimization methods

There are various optimization techniques available to numerically solve the optimization problem of equation (5.5). One mainly differentiates between local and global, and between gradient-based and derivative-free optimization methods [106]. The performance of each of these methods strongly depends on the underlying problem. We here introduce a local and gradient-based optimization approach efficiently estimating the model parameters of deterministic models $\mathcal{M}(\boldsymbol{\theta})$ as described in section 4.2 [91].

Sensitivity equations The general idea of gradient-based optimization is to exploit the information embedded in the derivative of the objective function $J(\boldsymbol{\theta})$ to determine an adequate search direction. A naive way of approximating the gradient of an objective function $J(\boldsymbol{\theta})$ can

be obtained by applying the finite difference method [91]. A more sophisticated and efficient approach is provided by the sensitivity equations [91]. Assuming an additive Gaussian noise model as described in equation (5.9), the partial derivative of the objective function with respect to parameter θ_l is given by

$$\begin{aligned}\frac{\partial}{\partial\theta_l}J(\boldsymbol{\theta}) =& \frac{1}{2}\sum_{k=1}^{n_t}\sum_{i=1}^{n_y}\frac{1}{\sigma_{i,k}^2(\boldsymbol{\theta})}\left(1-\left(\frac{y_i^{\mathcal{D}}(t_k)-y_i(t_k,\boldsymbol{\theta})}{\sigma_{i,k}(\boldsymbol{\theta})}\right)^2\right)\frac{\partial}{\partial\theta_l}\sigma_{i,k}^2(\boldsymbol{\theta}) \\ &- 2\frac{y_i^{\mathcal{D}}(t_k)-y_i(t_k,\boldsymbol{\theta})}{\sigma_{i,k}^2(\boldsymbol{\theta})}\frac{\partial}{\partial\theta_l}y_i(t_k,\boldsymbol{\theta}).\end{aligned}$$

Then the sensitivities, describing the partial derivatives of the observable and state vectors are defined as below.

Definition 5.3. *The sensitivity of the observable* $\mathbf{y}(t,\boldsymbol{\theta})$ *with respect to* θ_l *is given by*

$$\mathbf{s}_l^y(t) = \frac{\partial}{\partial\theta_l}\mathbf{y}(t,\boldsymbol{\theta}) \in \mathbb{R}^{n_y} \tag{5.13}$$

and the sensitivity of the state $\mathbf{x}(t,\boldsymbol{\theta})$ *with respect to* θ_l *is given by*

$$\mathbf{s}_l^x(t) = \frac{\partial}{\partial\theta_l}\mathbf{x}(t,\boldsymbol{\theta}) \in \mathbb{R}^{n_x}. \tag{5.14}$$

Then the sensitivity $\mathbf{s}_l^y(t)$ with respect to θ_l of the observable $\mathbf{y}(t,\boldsymbol{\theta})$ (equation (4.5)) is given by

$$\mathbf{s}_l^y(t) = \frac{\partial}{\partial\theta_l}h(t,\boldsymbol{\theta},\mathbf{x}(t,\boldsymbol{\theta})) + \frac{\partial}{\partial x}h(t,\boldsymbol{\theta},\mathbf{x}(t,\boldsymbol{\theta}))\,\mathbf{s}_l^x(t), \tag{5.15}$$

where $\frac{\partial}{\partial x}h(t,\boldsymbol{\theta},\mathbf{x}(t,\boldsymbol{\theta})) = \left(\frac{\partial}{\partial x_m}h_j(t,\boldsymbol{\theta},\mathbf{x}(t,\boldsymbol{\theta}))\right)_{mj}$, with

$$\begin{aligned}\frac{\partial}{\partial t}\mathbf{s}_l^x(t) &= \frac{\partial}{\partial x}f(t,\boldsymbol{\theta},\mathbf{x}(t,\boldsymbol{\theta}))\,\mathbf{s}_l^x(t) + \frac{\partial}{\partial\theta_l}f(t,\boldsymbol{\theta},\mathbf{x}(t,\boldsymbol{\theta})) \\ \mathbf{s}_l^x(0) &= \frac{\partial}{\partial\theta_l}\mathbf{x}_0.\end{aligned}$$

Hence, to compute the sensitivities $\mathbf{s}_l^y(t)$ one needs to compute the solution to an ODE system of size n_x for every θ_l in $\boldsymbol{\theta}$, where n_x the number of states of the model ODE [106]. Generally, $\mathbf{s}_l^y(t)$ can be efficiently and robustly computed [113]. The gradient of the negative log-likelihood for data with a Laplacian distributed noise can be derived similarly.

Multi-start optimization Generally, one would like to obtain the global optimum of the minimization problem described in equation (5.5). However, a common problem when estimating the parameters of a dynamic system is the existence of multiple local optima in the respective objective function $J(\boldsymbol{\theta})$. A single run of gradient descent might then converge to a local rather than a global optimum [91]. To make deterministic optimization approaches more robust, one can perform multiple independent optimization runs where each run is initiated from a different point in the parameter space Θ [91]. The optimum with the lowest negative log-likelihood value is used as an approximation to the global optimum of the minimization problem and its reliability increases with the number of optimization runs converging to the same local optimum.

5.6. Identifiability and uncertainty analysis

When estimating the parameter vector $\boldsymbol{\theta}$ of $\mathcal{M}(\boldsymbol{\theta})$ one is interested in identifying a unique and optimal solution. Whether an optimization problem as stated in equation (5.5) gives an unique

solution $\boldsymbol{\theta}^*$ depends on both the functions f and h of $\mathcal{M}(\boldsymbol{\theta})$ (see 4.2) and the experimental data $\mathcal{D}$ used to estimate $\boldsymbol{\theta}^*$. Thereby we differentiate between structural and practical identifiability [114].

Definition 5.4. *A parameter θ_i is structurally local identifiable if for any $\boldsymbol{\theta}' \in \Theta$ there exists a neighborhood $V(\boldsymbol{\theta}')$ such that*

$$\boldsymbol{\theta} \in V(\boldsymbol{\theta}') \text{ and } \mathcal{M}(\boldsymbol{\theta}) = \mathcal{M}(\boldsymbol{\theta}') \Rightarrow \theta_i = \theta_i'. \tag{5.16}$$

Structural identifiability only considers functions f and h of $\mathcal{M}(\boldsymbol{\theta})$ and answers the question whether a parameter θ_i can be uniquely identified given perfect experimental data $\mathcal{D}$, i.e., noise-free and containing infinite measurements [70]. Structural non-identifiability arises from a redundant parameterization due to an insufficient mapping h of the states $\mathbf{x}(t, \boldsymbol{\theta})$ to the observations $\mathbf{y}(t, \boldsymbol{\theta})$ [114]. To circumvent structural non-identifiability, measurements have to be taken in accordance with a sufficient mapping h. To test for structural identifiability one can either directly compare the vector fields of $f(t, \boldsymbol{\theta}, \mathbf{x}(t, \boldsymbol{\theta}))$ and $f(t, \boldsymbol{\theta}', \mathbf{x}(t, \boldsymbol{\theta}'))$ (direct approach), use approximation approaches such as the Taylor or generating series, or use available toolboxes such as GenSSI [115], STRIKE-GOLDD [116] or DAISY [117].

Example. We introduce an example, where the model parameters are not structurally identifiable given the dynamics, initial conditions and measurements [118]. Let the system dynamics be described by the following equations

$$\frac{\partial}{\partial t} x_1(t, \boldsymbol{\theta}) = -\theta_1 x_1$$
$$\frac{\partial}{\partial t} x_2(t, \boldsymbol{\theta}) = \theta_1 x_1$$

with initial conditions $x_1(t_0, \boldsymbol{\theta}) = \theta_2$, $x_2(t_0, \boldsymbol{\theta}) = 0$ and $y_1(t, \boldsymbol{\theta}) = \theta_3 x_2(t, \boldsymbol{\theta})$ the only measurement. Solving the system of ODEs and substituting the solution into the measurement equation gives

$$y_1(t, \boldsymbol{\theta}) = \theta_3 \theta_2 e^{-\theta_1 t} \left(e^{\theta_1 t} - 1 \right).$$

Here, θ_2 and θ_3 are structurally non-identifiable as one cannot differentiate between their product.

We further consider practical identifiability, where also the quality and abundance of the data measurements are taken into consideration.

Definition 5.5. *A $(1-\alpha)\%$ confidence interval/region of parameter θ_i is an observed interval/region $[L(\theta_i, \mathcal{D}), U(\theta_i, \mathcal{D})]$ for which*

$$\mathbb{P}(L(\theta_i, \mathcal{D}) \leq \theta_i \leq U(\theta_i, \mathcal{D})) = 1 - \alpha, \tag{5.17}$$

where α is called the significance level and $\mathcal{D}$ the data.

Definition 5.6. *A parameter θ_i is practically identifiable for a data set $\mathcal{D}$ if its $(1-\alpha)\%$ confidence interval $[L(\theta_i, \mathcal{D}), U(\theta_i, \mathcal{D})]$ is finite [114].*

Depending on the quality of data $\mathcal{D}$, a parameter θ_i which is structurally identifiable may still be practically unidentifiable [114]. To evaluate the confidence interval CI_i^α of parameters θ_i one can compute the profile likelihood for parameter θ_i by fixing θ_i and by optimizing over all other parameters.

Definition 5.7. *The profile likelihood for parameter θ_i is*

$$PL(\theta_i) = \underset{\substack{\theta_{j \neq i} \in \Theta \\ \text{subject to } \mathcal{M}(\theta)}}{\arg\max} \mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}). \tag{5.18}$$

Then the profile-likelihood-based confidence interval used for evaluating the practical identifiability of a parameter θ_i for significance level α is

$$CI_i^{\alpha} = \left\{ \theta_i | \exists \boldsymbol{\theta} : \frac{PL(\theta_i)}{\mathcal{L}_{\mathcal{D}}\left(\boldsymbol{\theta}^{ML}\right)} > e^{-\frac{\Delta_{\alpha}}{2}} \right\}, \tag{5.19}$$

where $\Delta_{\alpha} = \chi^2(\alpha, 1)$ for point-wise confidence and $\chi^2(\alpha, 1)$ the α-th percentile of the χ^2-distribution with 1 degree of freedom. The likelihood-ratio $\frac{PL(\theta_i)}{\mathcal{L}_{\mathcal{D}}\left(\boldsymbol{\theta}^{ML}\right)}$ is an indicator of the uncertainty of the estimate of parameter θ_i. In Schuh et al. [3], we used PESTO: Parameter EStimation TOolbox to evaluate the practical identifiability of the model parameters based on profile likelihoods [119].

6. Model selection

In the previous chapter, the model structure $\mathcal{M}(\boldsymbol{\theta})$ (see Definition 4.2) underlying the biological process of interest was assumed to be known. However, often there are several suggested and competing model hypotheses for which one would like to select the most plausible model(s) describing data set $\mathcal{D}$. Identifying the most plausible model(s) also presents an optimization problem, however, on an architectural level for which different optimization/selection methods are applied than for parameter estimation as introduced in chapter 5. Generally, model selection is only relative to the set of tested model hypotheses where models with a lower performance can be rejected but non-rejected models are not automatically validated [120]. The main philosophical principle underlying model selection is Occam's razor which states a preference for simple theories [121]. Hence, for model selection both the goodness of fit regarding the optimized model $\mathcal{M}(\boldsymbol{\theta}^*)$ given data $\mathcal{D}$ as well as the model's complexity are taken into consideration.

6.1. Bayes factor

Definition 6.1. *For two models $\mathcal{M}_i(\boldsymbol{\theta})$ and $\mathcal{M}_j(\boldsymbol{\theta}')$ the Bayes factor is defined as*

$$B_{ij} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_i(\boldsymbol{\theta}))}{\mathbb{P}(\mathcal{D}|\mathcal{M}_j(\boldsymbol{\theta}'))}, \tag{6.1}$$

with marginal likelihoods

$$\mathbb{P}(\mathcal{D}|\mathcal{M}_i(\boldsymbol{\theta})) = \int_{\Theta} \mathbb{P}(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_i(\boldsymbol{\theta}))\mathbb{P}(\boldsymbol{\theta}, \mathcal{M}_i(\boldsymbol{\theta}))d\boldsymbol{\theta}. \tag{6.2}$$

Adapted from [106].

The Bayes factor is a common criterion in model selection, where B_{ij} describes the relative support of model $\mathcal{M}_i(\boldsymbol{\theta})$ to model $\mathcal{M}_j(\boldsymbol{\theta}')$ [106]. By integrating over the parameter space Θ, the Bayes factor intrinsically accounts for increasing model complexity.

6.2. Bayesian Information Criterion

The computation cost for evaluating the marginal likelihoods is non-trivial for complex systems. Hence, one often reverts to an asymptotic approximation of the Bayes factor [122], as done in Schuh et al. [2, 3].

Definition 6.2. *The Bayesian Information Criterion (BIC) is defined as*

$$\mathrm{BIC}_i = n_\theta \log(n_y) - 2\log\left(\mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}^*)\right), \tag{6.3}$$

where n_y the number of observables in $\mathcal{D}$, n_θ the number of model parameters for model $\mathcal{M}_i(\boldsymbol{\theta})$ and $\mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}^*)$ the corresponding optimized likelihood function evaluated at $\boldsymbol{\theta}^*$ [123].

While the term $-2\log\left(\mathcal{L}_{\mathcal{D}}(\boldsymbol{\theta}^*)\right)$ evaluates the goodness of fit, the term $n_\theta \log(n_y)$ penalizes for the number of model parameters n_θ of model $\mathcal{M}_i(\boldsymbol{\theta})$. Hence, the lower the BIC value the better the overall model. For two models $\mathcal{M}_i(\boldsymbol{\theta})$ and $\mathcal{M}_j(\boldsymbol{\theta})$

$$\mathrm{BIC}_i - \mathrm{BIC}_j > 10, \tag{6.4}$$

evaluated on the same data set $\mathcal{D}$, then the model $\mathcal{M}_i(\boldsymbol{\theta})$ can be rejected [123]. Moreover, the BIC is only a valid model selection criterion for structurally identifiable models [106]. For a large

data set $\mathcal{D}$ and models $\mathcal{M}_i(\boldsymbol{\theta})$ and $\mathcal{M}_j(\boldsymbol{\theta})$ we have

$$\mathrm{BIC}_i - \mathrm{BIC}_j \approx 2\log(B_{ij}) \tag{6.5}$$

and the probability of the BIC selecting the true best model increases with increasing data set size [124].

6.3. Akaike Information Criterion

Similar to the BIC in form, another model selection criterion was derived from the principles of information theory.

Definition 6.3. *The Akaike Information Criterion (AIC) is given by*

$$\mathrm{AIC}_i = 2n_\theta - 2\log\left(\mathcal{L}_\mathcal{D}(\boldsymbol{\theta}^*)\right), \tag{6.6}$$

where n_θ the number of model parameters for model $\mathcal{M}_i(\boldsymbol{\theta})$ and $\mathcal{L}_\mathcal{D}(\boldsymbol{\theta}^*)$ the corresponding optimized likelihood function evaluated at $\boldsymbol{\theta}^*$ [125].

Relative to the AIC, the BIC penalizes additional model parameters or complexity more strongly. While the AIC and BIC are easily computed, they are still only approximations and can hence perform poorly in comparison to the Bayes factor. Other model selection criteria are given by but are not limited to the corrected AIC accounting for finite data sets [126] and the likelihood-ratio test which is only valid for nested models [127].

6.4. Application: External cue as an integral part of gene expression regulation

We give an application of a deterministic gene expression model, which makes use of the previously introduced concepts. We like to emphasize, that the application presented here is only briefly introduced and not all details are discussed. For a full description of this application see A.2. This section is based on, and is partly identical to, the core publication [2].

Research problem Among other mechanisms, the openness of DNA and hence gene expression is regulated by the placement of methyl groups on the tails of histone proteins. Particularly during development, where cells continuously divide and differentiate, a fast and economical control of gene expression is required. However, cells divide rapidly during development and with each cell cycle unmodified histones get incorporated into the DNA leading to an overall dilution of histone modifications. How is a precise and accurate gene expression assured during development despite the constant dilution of most histone modifications? In other words: Is the cell cycle as external cue an integral part of gene expression regulation?

Experiment and data To identify how the cell cycle shapes the histone modification landscape during development, we compared the kinetics of histone 4 lysine 20 methylation (H4K20me) in a population of normal frog embryos (from now on called 'mock') with a cell-cycle-arrested population. For this, half of the embryos were continuously incubated with hydroxyurea/aphidicolin (from now on called 'HUA') from 11 hours post fertilization onward, arresting the cell cycle. We measured the global distribution of un-, mono-, di-, and tri-methylated H4K20 by mass spectrometry at 14.75, 19.75, 27.5, and 40 hours post fertilization for three biological replicates (Figure 6.1).

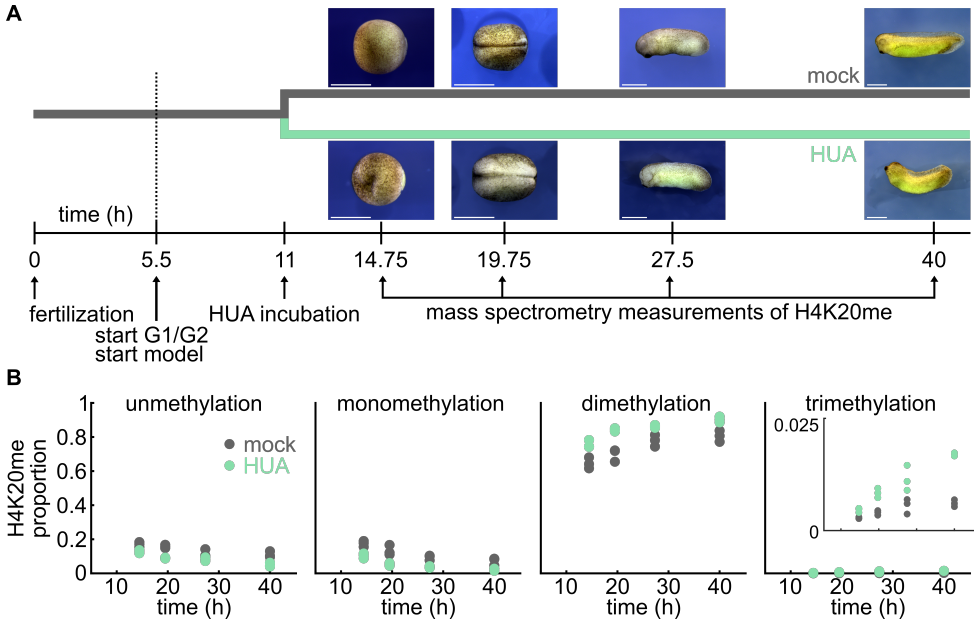


Figure 6.1: H4K20 methylation kinetics during frog development are altered upon HUA-induced cell cycle arrest. (A) Frog eggs are fertilized in vitro at time point 0. At 5.5 hours, G1 and G2 cell cycle phases appear. At 11 hours, half of the embryos are incubated with hydroxyurea/aphidicolin (HUA) arresting the cell cycle. Mass spectrometry measurements of H4K20 methylation (H4K20me) are performed at 14.75, 19.75, 27.5, and 40 hours in embryos with dividing (mock) or non-dividing cells (HUA). (B) H4K20me kinetics differ between mock (gray) and HUA treated (green) embryo populations. Adapted from [2].

Approach Effects of the cell cycle on H4K20me kinetics during frog development should lead to different H4K20me kinetics between mock and HUA. To identify whether we require population-specific rate constants to describe the H4K20me kinetics of mock and HUA, we developed a set of model hypotheses describing the H4K20me kinetics of mock and HUA. Model selection and parameter estimation on the joint model hypotheses allowed us to directly identify essential population-specific rate constants for mock and HUA.

Mock models To first identify the model structure underlying the average kinetics of H4K20me proportions of mock embryos only, we established a general mock model. This model will then give rise to a whole set of nested mock model hypotheses. We assumed that methylation, the addition of a methyl group, and demethylation, the active removal of a methyl group, occur successively. Then, for cells undergoing cell division, newly synthesized and unmethylated histones are incorporated into replicating DNA leading to an increase of unmethylated H4K20me. With cell division, the number of cells increases with the population growth rate

$$g(t) = \frac{\ln(2)}{c(t)}, \tag{6.7}$$

where $c(t)$ is the average cell cycle duration across all cells of an average mock embryo as a function of time. We tested five different cell cycle functions $c(t)$ describing the average cell cycle duration, where $c_1(t) = a$ assumes a constant cell cycle duration, $c_2(t) = a + bt$ assumes a linearly increasing cell cycle duration, $c_3(t) = 0.5 + bt$ assumes a linearly increasing cell cycle duration with a cell cycle duration of half an hour at 5.5 hours post fertilization (according to

prior experimental information), $c_4(t) = a + b\frac{t}{h+t}$ assumes the cell cycle duration to follow a scaled Hill function with Hill coefficient 1 and offset a, $c_5(t) = 0.5 + b\frac{t}{h+t}$ assumes the cell cycle duration to follow a scaled Hill function with Hill coefficient 1 and offset 0.5, or $c_6(t) = 0.5 + b\frac{t}{b+t}$ assumes the cell cycle duration to follow a constrained scaled Hill function with Hill coefficient 1 and offset 0.5.

The model is initiated at 5.5 hours post fertilization, as G1 and G2 cell cycle phases are only established then [128]. At model initialization, experimental data suggests an average cell cycle duration of half an hour, justifying cell cycle functions $c_3(t)$, $c_5(t)$ and $c_6(t)$ with fixed offset of 0.5 hours [129, 130]. Furthermore, cell cycle functions c_4, c_5 and c_6 make use of the non-linearity of the Hill function (section 4.3). Let $\mathcal{C}_{abs} = \{\text{me}i_{abs}\}_{i=0}^{3}$ be the set of chemical species describing absolute un- (me0), mono- (me1), di- (me2) and tri-methylated (me3) H4K20, where $X_{\text{me}i}$ is the absolute H4K20 methylation as measured by mass spectrometry for chemical species me$i$. For the absolute mock H4K20me, we formulated a chemical reaction network with reactions

$$
\begin{aligned}
R_1 &: \quad \text{me0} \xrightarrow{m_1} \text{me1} \\
R_2 &: \quad \text{me1} \xrightarrow{m_2} \text{me2} \\
R_3 &: \quad \text{me2} \xrightarrow{m_3} \text{me3} \\
R_4 &: \quad \text{me1} \xrightarrow{d_1} \text{me0} \\
R_5 &: \quad \text{me2} \xrightarrow{d_2} \text{me1} \\
R_6 &: \quad \text{me3} \xrightarrow{d_3} \text{me2} \\
R_7 &: \quad \text{me0} \xrightarrow{g(t)} 2\text{me0} \\
R_8 &: \quad \text{me1} \xrightarrow{g(t)} \text{me1} + \text{me0} \\
R_9 &: \quad \text{me2} \xrightarrow{g(t)} \text{me2} + \text{me0} \\
R_{10} &: \quad \text{me3} \xrightarrow{g(t)} \text{me3} + \text{me0}
\end{aligned}
$$

where, m_1, m_2, m_3 methylation rate constants, d_1, d_2, d_3 demethylation rate constants and $g(t)$ the population growth rate (Figure 6.2). In reactions R_7, R_8, R_9, and R_{10} we describe the DNA replication before cell division, where the new integrated histones are unmethylated.

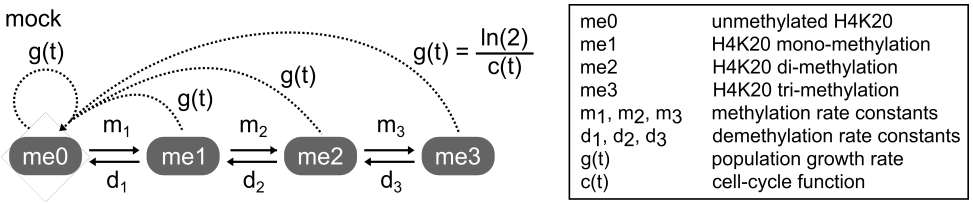


Figure 6.2: Model of cycling mock population composed of four H4K20 states: un- (me0), mono- (me1), di- (me2), and tri-methylation (me3). m_1, m_2, and m_3 represent the mono-, di-, and tri-methylation rate constants and d_1, d_2, and d_3 represent the demethylation rate constants. Due to cell division there is an overall increase of H4K20me0 parameterized with population growth rate g(t), which is dependent on the cell cycle function c(t). The dotted arrows represent reactions without conversions. Adapted from [2].

As we considered the mean absolute H4K20me of a large number of histones across cells and

even across frog embryos, we assumed a deterministic modeling approach to be appropriate. The system of ODEs describing the mock model for mean absolute H4K20me according to equation (4.1) is given by

$$\begin{aligned}
\frac{d}{dt}X_{\mathrm{me0}}(t) &= -m_1 X_{\mathrm{me0}}(t) + d_1 X_{\mathrm{me1}}(t) + g(t)\sum_{i=0}^{3} X_{\mathrm{mei}}(t)\\
\frac{d}{dt}X_{\mathrm{me1}}(t) &= m_1 X_{\mathrm{me0}}(t) - (m_2 + d_1)X_{\mathrm{me1}}(t) + d_2 X_{\mathrm{me2}}(t)\\
\frac{d}{dt}X_{\mathrm{me2}}(t) &= m_2 X_{\mathrm{me1}}(t) - (m_3 + d_2)X_{\mathrm{me2}}(t) + d_3 X_{\mathrm{me3}}(t)\\
\frac{d}{dt}X_{\mathrm{me3}}(t) &= m_3 X_{\mathrm{me2}}(t) - d_3 X_{\mathrm{me3}}(t)\\
\frac{d}{dt}N(t) &= g(t)N(t),
\end{aligned}$$

where $N(t)$ is the total number of histone tails at time t. The experimental H4K20 methylation data is taken from different mock embryo populations for each of the four time points. To account for the varying absolute numbers in embryos, and hence varying absolute numbers in histone tails and H4K20me sites, we considered the proportions of H4K20me. We reformulated the system of ODEs for $Y_{\mathrm{mei}}(t) = \frac{X_{\mathrm{mei}}(t)}{N(t)}$, for i $\in \{0,1,2,3\}$ according to [4]. The corresponding ODEs are given by

$$\frac{d}{dt}Y_{\mathrm{me}i}(t) = \frac{\frac{d}{dt}X_{\mathrm{me}i}(t)}{N(t)} - \frac{X_{\mathrm{me}i}(t)N(t)}{\frac{d}{dt}N(t)}, \tag{6.8}$$

more specifically by

$$\begin{aligned}
\frac{d}{dt}Y_{\mathrm{me0}}(t) &= -m_1 Y_{\mathrm{me0}}(t) + d_1 Y_{\mathrm{me1}}(t) + g(t)\sum_{i=0}^{3} X_{\mathrm{mei}}(t)\\
\frac{d}{dt}Y_{\mathrm{me1}}(t) &= m_1 Y_{\mathrm{me0}}(t) - (m_2 + d_1 + g(t))\, Y_{\mathrm{me1}}(t) + d_2 Y_{\mathrm{me2}}(t)\\
\frac{d}{dt}Y_{\mathrm{me2}}(t) &= m_2 Y_{\mathrm{me1}}(t) - (m_3 + d_2 + g(t))\, Y_{\mathrm{me2}}(t) + d_3 Y_{\mathrm{me3}}(t)\\
\frac{d}{dt}Y_{\mathrm{me3}}(t) &= m_3 Y_{\mathrm{me2}}(t) - (d_3 + g(t))\, Y_{\mathrm{me3}}(t)\\
\frac{d}{dt}N(t) &= g(t)N(t),
\end{aligned}$$

such that cell division leads to an increase in $Y_{\mathrm{me0}}(t)$ and a decrease in $Y_{\mathrm{mei}}(t)$ for $i \in \{1,2,3\}$. The resulting model for $Y_{\mathrm{me}i}$, $\forall i \in \{0,1,2,3\}$, is described by Figure 6.2, where the non-conversion reactions, denoted by the dotted arrows, are changed to conversion reactions, such that there is an additional constant outflow of $Y_{\mathrm{me}i}$ for $i = \{0,1,2,3\}$ to Y_{me0}.

HUA models Similar to mock, we first formulated a general HUA model to identify the specific rate constants necessary to describe the average kinetics of H4K20me proportions of HUA embryos. The general HUA model will then be used to give rise to a set of HUA model hypotheses. Again we assumed that methylation and demethylation occur successively. Furthermore, we assumed that the HUA treatment at 11 hours post fertilization acts immediately and arrests the cell cycle of all cells instantaneously. According to these model assumptions, we first formulated a chemical reaction network for chemical species $\mathcal{C} = \{\mathrm{me}i\}_{i=0}^{3}$ describing the mean absolute HUA

H4K20me kinetics, with reactions

$$
\begin{aligned}
R_1: &\quad \text{me0} \xrightarrow{m_1} \text{me1} \\
R_2: &\quad \text{me1} \xrightarrow{m_2} \text{me2} \\
R_3: &\quad \text{me2} \xrightarrow{m_3} \text{me3} \\
R_4: &\quad \text{me1} \xrightarrow{d_1} \text{me0} \\
R_5: &\quad \text{me2} \xrightarrow{d_2} \text{me1} \\
R_6: &\quad \text{me3} \xrightarrow{d_3} \text{me2}
\end{aligned}
$$

where, m_1, m_2, m_3 the methylation rate constants and d_1, d_2, d_3 the demethylation rate constants (Figure 6.3).

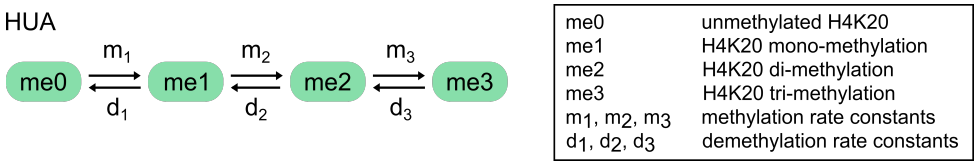


Figure 6.3: Model of cell-cycle-arrested HUA population. In contrast to the mock model (Figure 6.2), the HUA cells do not divide ($g(t) = 0$), and no dilution of methylated H4K20 is required. Adapted from [2].

We again assume a deterministic modeling approach to be appropriate for modeling the kinetics of mean absolute HUA H4K20me. The system of ODEs describing the mean absolute HUA H4K20me according to equation (4.1) is given by

$$
\begin{aligned}
\frac{d}{dt}X_{\text{me0}}(t) &= -m_1 X_{\text{me0}}(t) + d_1 X_{\text{me1}}(t) \\
\frac{d}{dt}X_{\text{me1}}(t) &= m_1 X_{\text{me0}}(t) - (m_2 + d_1) X_{\text{me1}}(t) + d_2 X_{\text{me2}}(t) \\
\frac{d}{dt}X_{\text{me2}}(t) &= m_2 X_{\text{me1}}(t) - (m_3 + d_2) X_{\text{me2}}(t) + d_3 X_{\text{me3}}(t) \\
\frac{d}{dt}X_{\text{me3}}(t) &= m_3 X_{\text{me2}}(t) - d_3 X_{\text{me3}}(t) \\
\frac{d}{dt}N(t) &= 0,
\end{aligned}
$$

where $X_{\text{me}i}$ is the absolute HUA H4K20me as measured by mass spectrometry for chemical species me$i$ and $N(t)$ is the total number of histone tails. We again reformulated the system of ODEs of absolute H4K20me to proportions according to $Y_{\text{mei}}(t) = \frac{X_{\text{mei}}(t)}{N(t)}$, i $\in \{0, 1, 2, 3\}$. However, for the HUA model the system of ODEs is identical for $X_{\text{mei}}(t)$ and $Y_{\text{mei}}(t)$, $\forall i = \{0, 1, 2, 3\}$.

Analysis and results of mock and HUA models In the previous two sections, we introduced the most general mock and HUA models. In addition, we tested several nested models for bot mock and HUA, such that rate constants could be shared between two or more reactions (termed 'shared methylation/demethylation rate constants') or be specific to one reaction (termed 'specific methylation/demethylation rate constants') (Figure 6.4). Moreover, we allowed for models without demethylation, as it has been postulated that demethylation might not be important for methylation kinetics at all [131].

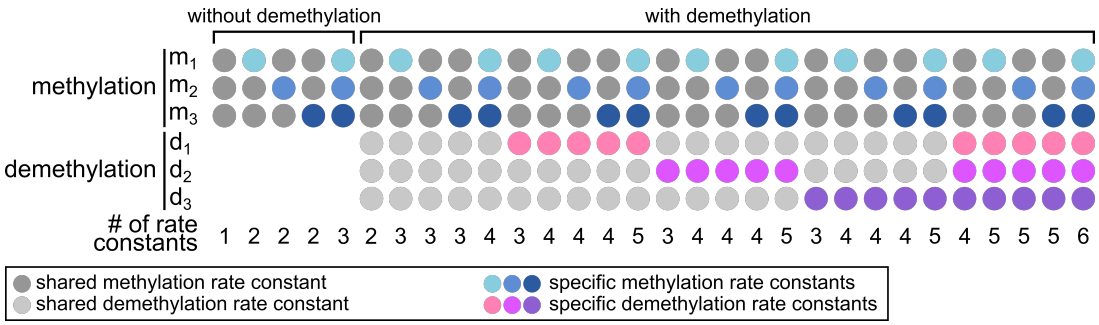


Figure 6.4: All possible rate constant combinations result in 5 models without demethylation and 25 models with demethylation. Rate constants specific to a particular methylation or demethylation step are indicated in color, rate constants shared between methylation or demethylation steps are shown in gray. The number of rate constants ranges between 1 and 6. Adapted from [2].

In total, we tested 180 mock model hypotheses (30 model hypotheses times 6 different cell cycle functions) and 30 HUA model hypotheses. We performed multi-start maximum likelihood optimization and model selection according to the Bayesian Information Criterion (BIC, see 6.2). We found that

(i) 12 out of 180 mock model hypotheses and 5 out of 30 HUA model hypotheses considerably outperformed the other model hypotheses with a BIC difference > 10.

(ii) the best performing models of mock and HUA show a good overall agreement with the respective H4K20me proportion kinetics.

(iii) only a constrained scaled Hill function with Hill coefficient 1 and offset 0.5, $c_6(t) = 0.5 + b\frac{t}{b+t}$ was able to correctly predict biologically meaningful average cell cycle duration of around 8 hours for the mock embryo population.

(iv) specific mono-, di, and tri-methylation rate constants or only a specific tri-methylation rate constant are required to describe the mock H4K20me proportion kinetics.

(v) demethylation rate constants do not need to be specific in the mock model and demethylation is not required to capture the mock H4K20me proportion kinetics.

(vi) specific mono-, di, and tri-methylation rate constants are required to describe the HUA H4K20me proportion kinetics.

(vii) demethylation rate constants do not need to be specific in the HUA model but demethylation is required to capture the HUA H4K20me proportion kinetics.

Joint models To identify which rate constants are mock and HUA population-specific, we formulated a general joint model based on the previous findings describing the H4K20me proportions for mock and HUA simultaneously. Similar to mock and HUA, the general joint model gives rise to a set of joint model hypotheses. For the joint model we assumed that mock and HUA H4K20me proportion kinetics have the same underlying model structure with regard to the specific and shared methylation and demethylation rate constants. The general joint model

is a combination of the best and simplest mock and HUA models (Figure 6.5). Its chemical reaction network and ODE system is identical to the combined formulations and ODE systems of mock and HUA as introduced above. As a constrained scaled Hill function with Hill coefficient 1 and offset 0.5 was the only function that led to biologically meaningful cell cycle duration, we only considered this cell cycle function for the mock population of the joint model. Furthermore, we allowed for 3 mock-specific and 3 HUA-specific methylation rate constants in the joint model, while we restricted demethylation to the simplest case of at most one shared mock-specific and one shared HUA-specific demethylation rate constant.

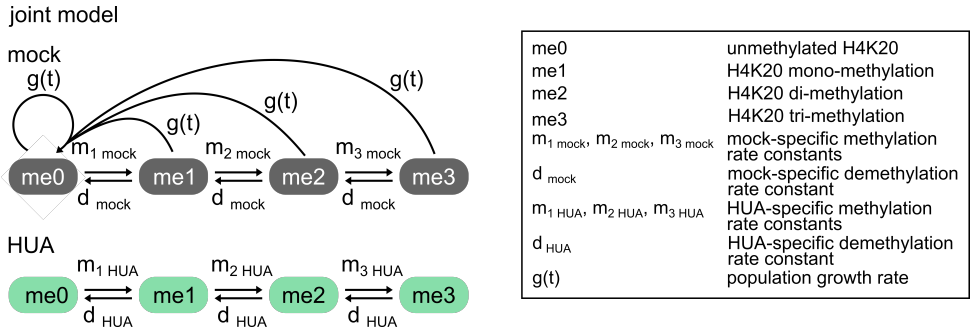


Figure 6.5: Joint model allows for three methylation and one demethylation rate constants for both mock and HUA as suggested by the best models for mock and HUA. Adapted from [2].

Analysis and results of joint model In addition to the general joint model (Figure 6.5), we tested several nested models with joint methylation and demethylation rate constants between mock and HUA and nested models without demethylation rate constants to determine which rate constants are substantially different between the two embryo populations (Figure 6.6).

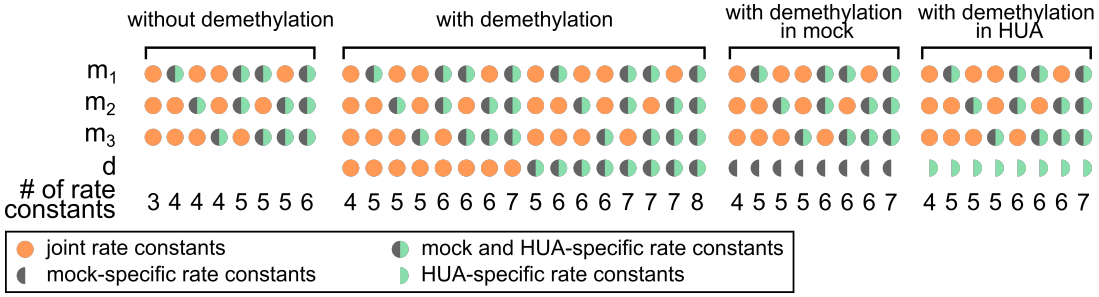


Figure 6.6: We fit 16 models with demethylation and 8 models each without demethylation in mock and/or HUA to the joint data to infer mock- and HUA-specific rate constants. The joint rate constants of mock and HUA are shown in orange, the rate constants present in both the mock and HUA models but taking on mock- and HUA-specific values are indicated in gray/green, the rate constants only present in the mock or HUA model are shown in gray and green half-circles, respectively. The model structure of the most complex of model is shown in Figure 6.5. The number of rate constants ranges between 3 and 8. Adapted from [2].

To identify joint model hypotheses that are able to explain the kinetics of H4K20me proportions in both mock and HUA, we fitted all 40 joint models using multi-start maximum likelihood optimization and performed model selection according to the BIC (see 6.2). We found that

(i) 6 out of the 40 joint model hypotheses considerably outperformed the other model hypotheses.

(ii) the best performing joint models showed a good overall agreement with mock and HUA H4K20me proportion kinetics (Figure 6.7).

(iii) mono- and di-methylation rate constants are required to be mock- and HUA-specific.

(iv) tri-methylation may be described by a joint rate constant.

(v) demethylation is essential to describe the kinetics of H4K20me proportions for HUA but not necessary for mock.

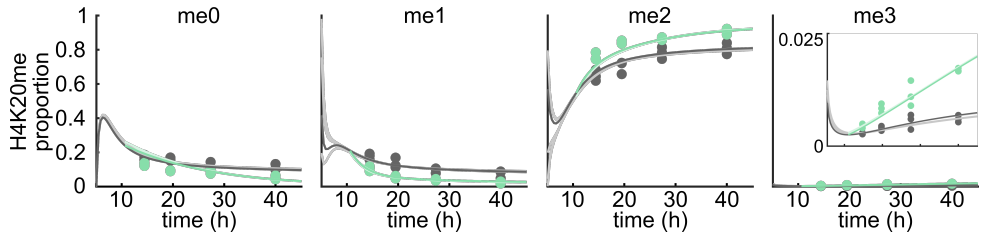


Figure 6.7: All best-performing joint models are able to explain both the mock and HUA data. The estimated initial conditions vary between the models. The joint model with the lowest BIC is highlighted. Adapted from [2].

Conclusion In sum, we developed a deterministic mathematical model describing the joint kinetics of H4K20me proportions to identify rate constants specific to mock- and HUA embryo populations. Our approach revealed that cell cycle plays an active role in shaping the global H4K20me landscape during development. Moreover, it revealed that demethylation is essential for regulating H4K20me kinetics in HUA cells, while it is likely dispensable in mock cells. This suggests that cell-cycle-mediated dilution of H4K20 methylation is an essential regulatory component for shaping the H4K20me landscape during early development. Together, we show that an external cue can be an integral part of gene expression regulation.

6.5. Application: Repeated external cue and adaptive gene expression regulation of repression

We give a second application of a deterministic gene expression model, which makes use of the previously introduced concepts. We like to emphasize, that the application presented here is only briefly introduced and not all details are discussed. For a full description of this application see B.1. This section is based on, and is partly identical to, the manuscript under review [3].

Research problem Adaptation to changing environments is crucial for cell survival. When exposed to repeated stimulation cells can remember their initial transcriptional response which leads to an adapted reaction upon re-stimulation. This phenomenon is termed transcriptional memory. Most studies on transcriptional memory so far have focused on gene induction, i.e., the activation of genes, although gene repression, i.e., the deactivation of genes, plays a similarly central role in gene regulation. This opens the important questions as to whether memory also exists for repression and how repression memory manifests at a single-cell level. More generally speaking: Can a repeated external cue result in an adaptive gene expression regulation of repression?

Experiment and data To identify whether budding yeast cells show memory in repeated repressions of the gene *Galactokinase 1* (Gal1), we measured Gal1 expression via a Gal1-GFP (green fluorescent protein) fusion in budding yeast cells across repeated galactose inductions and glucose repressions using time-lapse microscopy coupled to a microfluidics device (see section 1.1 and Figure 6.8A). Gal1 converts galactose into energy and is hence only expressed when yeast cells are exposed to medium containing galactose. Images of each microfluidics chamber were taken every 3 minutes totaling to 320 images per chamber during one full 16-hour experiment. Yeast cells were then semi-automatically segmented, mapped and the total Gal1-GFP fluorescence signal, per cell and per time point, was extracted using PhyloCell and Autotrack resulting in over 700 single-cell Gal1 expression traces (Figure 6.8B). In contrast to other cells, asymmetric budding of yeast cells allowed for lineage tracking such that mother-daughter relationships were identified. We focused our analysis on the repression kinetics of the first two hours of repression.

Approach We assumed that transcriptional memory of repression would lead to an adaptive dynamic response of Gal1 in the repeated repression. To identify whether the Gal1 repression kinetics indeed differ between repressions r1 and r2 in budding yeast cells, we formulated a mathematical framework to describe the repression kinetics of single-cell Gal1 traces specific to each repression, and performed parameter estimation allowing for a systematic comparison of model parameters between repressions r1 and r2.

Dilution compensation In glucose, yeast cells divide rapidly. During the emergence of a new daughter cell the Gal1 proteins are distributed across the mother and daughter cell. Assuming stable total Gal1 protein amount, the redistribution leads to a decrease of total Gal1-GFP fluorescence signal in the mother cell (Figure 6.8C top), an effect called dilution. To deconvolute the dilution and repression kinetics in the total Gal1-GFP fluorescence signal and to compensate for GFP dilution, we created an artificial population of quasi non-dividing cells by summing up the total Gal1-GFP fluorescence signal of the mother cell, i.e., a cell existing at the start of a respective repression, and its progeny (Figure 6.8C bottom). The total GFP of an artificially created non-dividing cell is defined as

$$\text{total GFP}(t) = \text{Gal1-GFP}_{\text{mother}}(t) + \sum_{i=1}^{N} \text{Gal1-GFP}_{\text{progeny } i}(t) \quad \forall t \in \mathbb{R}, \tag{6.9}$$

where $\text{Gal1-GFP}_{\text{mother}}(t)$ is the Gal1-GFP fluorescence value of the mother cell at time t and $\text{Gal1-GFP}_{\text{progeny } i}(t)$ is the Gal1-GFP fluorescence value of progeny cell i at time t, assuming the mother cell to give rise to N progeny cells in the respective repression. If the mother cell is connected to a daughter cell at the start of a repression, the daughter cell counts as a progeny cell. We applied this method of dilution compensation to the single-cell Gal1 traces of repressions r1 and r2 to deconvolute the dilution and repression kinetics in this study (Figure 6.8D).

Repressor and non-repressor model As galactose induction of Gal1 is highly variable with respect to its induction delay at the single-cell level, only a subpopulation of cells induce Gal1 during galactose induction [132]. Only cells which induced Gal1 during galactose induction are assumed to show repression kinetics during glucose repression. To describe the repression kinetics of induced and non-induced single budding yeast cells we formulated two mathematical

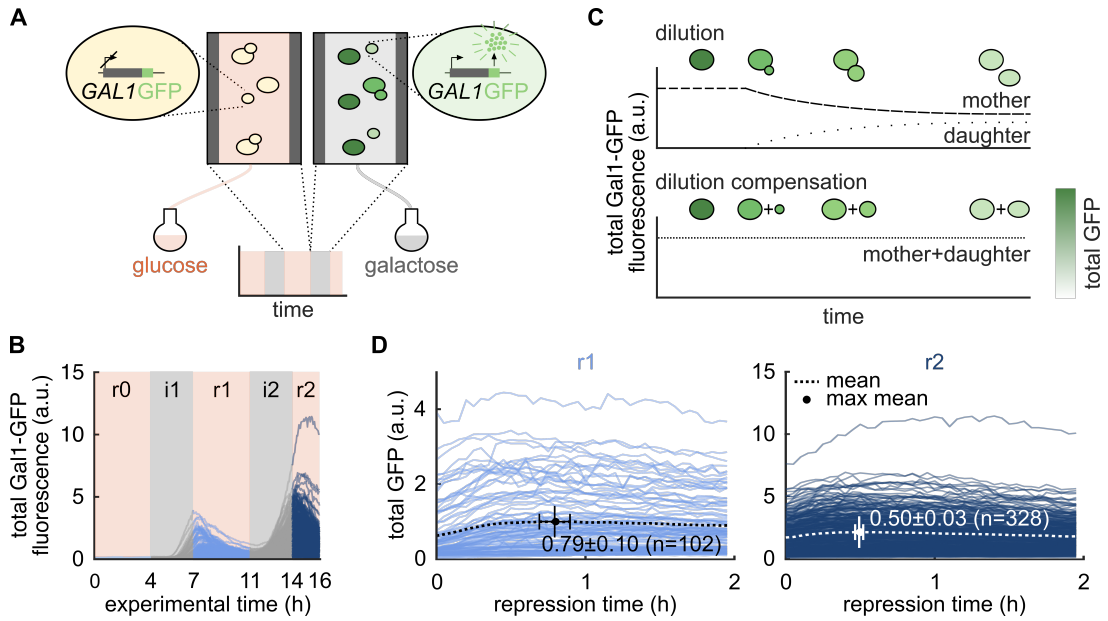


Figure 6.8: Gal1 repression experiment and data. (A) Budding yeast cells are grown in microfluidics chambers and alternatingly exposed to medium containing either glucose (orange) or galactose (gray) as sugar source. Gal1 is induced in cells exposed to galactose and repressed in cells exposed to glucose. Gal1 expression was monitored via a Gal1-GFP fusion and time-lapse microscopy. (B) Single-cell traces of total Gal1-GFP fluorescence signal in arbitrary units (a.u.) of budding yeast cells across two inductions (gray) and repressions 0, 1 and 2 (blue). (C) For stable Gal1 expression, budding leads to a decrease of total Gal1-GFP fluorescence signal in the mother cell and to an increase of total Gal1-GFP fluorescence signal in the daughter cell (top). To compensate for this dilution, we added up the total Gal1-GFP fluorescence signal of each mother cell present at the beginning of a glucose repression period and its progeny (bottom). (D) Single-cell traces of total GFP signal compensated for dilution for the first two hours of repression 1 (r1, left) and repression 2 (r2, right). The mean expressions are highlighted by dotted lines and the maximal mean total GFP is highlighted by a dot. Taken from [3].

models. As Gal1 induction leads to an approximate 1000-fold change in Gal1 expression [133], stochasticity inherent to gene expression was assumed to be negligible. Hence, a deterministic modeling approach is sufficient to explain the kinetics of the total GFP over time per cell. We further assumed that all reactions resulting in the production of GFP can be summarized by a constant reaction propensity. Under these model assumptions, we defined a chemical reaction network describing the production and degradation of total GFP for a non-repressor cell with chemical specie $\mathcal{C} = \{\text{GFP}\}$, the total GFP, and reactions

$$R_1: \quad \emptyset \xrightarrow{r_{\text{basal}}} \text{GFP}$$
$$R_2: \quad \text{GFP} \xrightarrow{r_{\text{deg}}} \emptyset,$$

where r_{basal} is the basal GFP production rate constant and r_{deg} is the GFP degradation rate constant (Figure 6.9 left). The ODE describing the non-repressor model is given by

$$\frac{\partial}{\partial t}\text{GFP}(t) = r_{\text{basal}} - r_{\text{deg}}\text{GFP}(t), \tag{6.10}$$

with the solution

$$\text{GFP}(t) = \frac{r_{\text{basal}}}{r_{\text{deg}}} + (\text{GFP}_0 - \frac{r_{\text{basal}}}{r_{\text{deg}}})e^{-r_{\text{deg}}t} = \frac{r_{\text{basal}}}{r_{\text{deg}}}(1 - e^{-r_{\text{deg}}t}) + \text{GFP}_0 e^{-r_{\text{deg}}t}, \tag{6.11}$$

where $\mathrm{GFP}_0 = \mathrm{GFP}(t_0)$, the initial total GFP at time point t_0. The non-repressor model comprises 4 model parameters, such that $\boldsymbol{\theta}_{\text{non-repressor}\,i} = \{\mathrm{GFP}_{0\,i}, r_{\text{basal}\,i}, r_{\deg\,i}, \sigma_i\}$ for cell i, where σ_i the noise parameter determining the width of the Gaussian noise distribution (see 5.4).

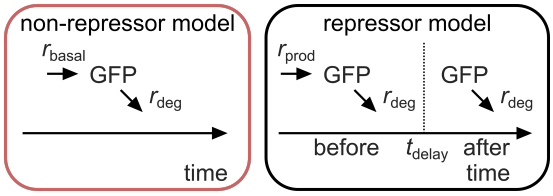


Figure 6.9: Left: Model for non-repressor cells composed of basal GFP production (r_{basal}) and degradation ($r_{\deg}$). Right: Model for repressor cells composed of an initial constant GFP production (r_{prod}) and degradation ($r_{\deg}$) until a delayed repression onset t_{delay} where GFP production is switched off. Taken from [3].

For cells showing repression kinetics we assumed that the termination of Gal1 expression is delayed but instantaneous and that GFP is expressed fast, such that GFP kinetics reflect the active and inactive states of the Gal1 gene. Under these additional model assumptions, we defined a chemical reaction network describing the GFP kinetics of repressor cells containing chemical specie $\mathcal{C} = \{\mathrm{GFP}\}$, the total GFP, and reactions

$$\begin{aligned}
&\underline{\text{before } t = t_{\text{delay}}}\\
&R_1: \qquad \emptyset \xrightarrow{r_{\text{prod}}} \mathrm{GFP}\\
&R_2: \qquad \mathrm{GFP} \xrightarrow{r_{\deg}} \emptyset\\
&\underline{\text{after } t = t_{\text{delay}}}\\
&R_1: \qquad \mathrm{GFP} \xrightarrow{r_{\deg}} \emptyset,
\end{aligned}$$

where r_{prod} the GFP production rate constant, $r_{\deg}$ the GFP degradation rate constant and t_{delay} the repression delay, i.e., the time between the carbon source switch form galactose to glucose and the termination of Gal1 expression (Figure 6.9 right). The ODE describing the repressor model before t_{delay} is given by

$$\frac{\partial}{\partial t}\mathrm{GFP}(t) = r_{\text{prod}} - r_{\deg}\mathrm{GFP}(t), \tag{6.12}$$

with the solution

$$\mathrm{GFP}(t) = \frac{r_{\text{prod}}}{r_{\deg}} + (\mathrm{GFP}_0 - \frac{r_{\text{prod}}}{r_{\deg}})e^{-r_{\deg}t} = \frac{r_{\text{prod}}}{r_{\deg}}(1 - e^{-r_{\deg}t}) + \mathrm{GFP}_0 e^{-r_{\deg}t}, \tag{6.13}$$

where $\mathrm{GFP}_0 = \mathrm{GFP}(t_0)$, the initial total GFP at time point t_0. The ODE describing the repressor model after t_{delay} is given by

$$\frac{\partial}{\partial t}\mathrm{GFP}(t) = -r_{\deg}\mathrm{GFP}(t), \tag{6.14}$$

with solution

$$\mathrm{GFP}(t) = \mathrm{GFP}(t_{\text{delay}})e^{-r_{\deg}(t - t_{\text{delay}})}, \tag{6.15}$$

where $\mathrm{GFP}(t_{\text{delay}}) = \frac{r_{\text{prod}}}{r_{\deg}}(1 - e^{-r_{\deg}t_{\text{delay}}}) + \mathrm{GFP}_0 e^{-r_{\deg}t_{\text{delay}}}$, the GFP at time point t_{delay}. The repressor model comprises 5 model parameters: $\boldsymbol{\theta}_{\text{repressor}\,i} = \{\mathrm{GFP}_{0\,i}, t_{\text{delay}\,i}, r_{\text{prod}\,i}, r_{\deg\,i}, \sigma_i\}$ for cell i, where σ_i the noise parameter determining the width of the Gaussian noise distribution (see 5.4). For $t_{\text{delay}} \geq 2$ the repressor model corresponds to the non-repressor model.

Analysis and results

(i) For each Gal1 trace in repression r1 and r2, we performed multi-start maximum likelihood optimization for the non-repressor and repressor model to infer the cell-specific and model-specific parameters (Figure 6.10) and performed model selection according to the Bayesian Information Criterion (BIC) (see Definition 6.2) to distinguish between non-repressor and repressor cells.

(ii) Using the estimated single-cell parameters of the repressor cell population, we found the repression delay to be significantly shortened upon the repeated repression, suggesting repression memory.

(iii) We repeated this analysis for a mutant yeast strain, *elp6*Δ, where the median repression delay is again shortened in the repeated repression, suggesting repression memory also in the mutant strain *elp6*Δ.

(iv) Comparing the estimated single-cell parameters of wildtype and *elp6*Δ repressor cells show that *elp6*Δ exhibits a stronger decrease in the repression delay between repressions r1 and r2, suggesting that *elp6*Δ is a gain-of-repression-memory mutant.

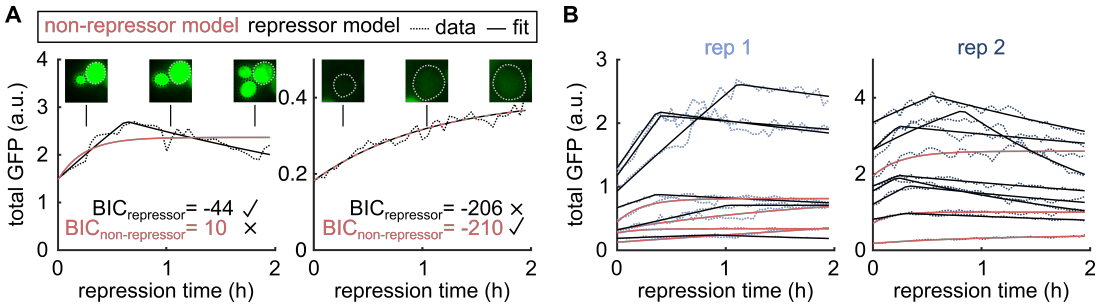


Figure 6.10: Model selection to discriminate between repressor and non-repressor cells. (A) Two total GFP traces (dotted line) and fits of the non-repressor model (red solid line) and repressor model (black solid line). Exemplary images of the cell(s) at three different time points are shown above. The mother cell is circled in gray and the progeny is circled in black. The better fitting model was selected according to the BIC. Left: total GFP trace better fitted by the repressor model. Right: total GFP trace fitted equally well by the non- repressor and repressor model. Due to the higher model complexity of the repressor model, the repressor model is rejected. (B) 10 exemplary total GFP traces (dotted lines) and best fits (solid lines) of cells for repressions 1 (left) and 2 (right). Fits of total GFP traces best fitted with a repressor model are shown in black and fits of total GFP traces best fitted with a non- repressor model are shown in red for repressions rep1 (left) and rep 2 (right). Adapted from [3].

Conclusion We developed a deterministic gene expression model to distinguish between repressor and non-repressor cells and to quantitatively describe single-cell repression kinetics. Using the estimated parameters of the repressor cells, we found that there is repression memory in yeast cells and that *elp6*Δ is a novel gain-of-repression-memory mutant. Together, we show that a repeated external cue can result in adaptive gene expression regulation for repression.

7. Summary and outlook

In this thesis, we dissected mechanisms of gene expression regulation by mathematically modeling the effects of intrinsic noise and external cues on gene expression. By developing a stochastic model comprising established principles of gene expression regulation, we found that intrinsic noise can result in rare and transient gene expression, qualitatively similar to the gene expression patterns observed in pre-resistant melanoma cells [1]. Next, we asked whether the cell cycle as a recurrent external cue can be an integral part of gene expression regulation. To describe global gene expression kinetics in frog embryos, where the cells are subject to rapid cell divisions, we developed deterministic models and performed parameter estimation and model selection. This analysis showed that the cell cycle is an integral regulatory component of global gene expression during frog development [2]. Finally, we investigated whether a repeated exposure to an external cue can lead to adaptive gene expression regulation of repression. By deterministically modeling and quantitatively comparing the gene expressions of Gal1 in single yeast cells, which were exposed to repeated carbon-source shifts, we indeed identified an adaptive gene expression regulation of Gal1 repression, i.e., the time to Gal1 expression termination is shortened upon repeated exposure to the external cue [3]. Our results provide new insights into and allow for a more comprehensive picture of the mechanisms of gene expression regulation.

The field of systems biology is rapidly advancing. This is highlighted by the continued identification of molecules and chemical modifications involved in gene expression regulation, such as tRNAs [134], rRNAs [135], as well as histone and mRNA modifications [136, 137]. This is partially facilitated by advances in experimental methods, such as mRNA-tagging for real-time imaging of translation [138] or scRNA-seq [14], and mathematical and computational methods for their analyses. We here outline future research directions for dissecting the mechanisms of gene expression regulation, restricting our outlook to the two, in our opinion, most promising mathematical and computational developments of the past years.

Until now, the quantification of single-cell gene expression levels, such as counting smRNA FISH spots [15] or segmenting, mapping and tracking cells to determine the fluorescence signals of GFP-reporters over time [67], was coupled to time-consuming, labor-intense manual annotations. This bottleneck limited the size of the final quantitative information extracted from images or videos. Deep learning revolutionized image processing in the past years [139]. Today, image analysis pipelines, such as CellPose [140] or YeaZ [141] allow for a fully automatized and reproducible image quantification, generating data almost comparable to human-performance in quality, but at a higher rate. More user-friendly, open-source software allows for a larger community of biomedical researchers to benefit from this new era of image processing [142, 143]. While not perfect yet, large image data sets will be easily, quickly, and reliably analyzed by computational image processing in the near future, allowing for the study of gene expression regulation via image data of thousands or millions of single cells. The mechanistic models developed throughout this thesis are not compromised by the size of the input data. Hence, no changes or adaptations are required to leverage more information of the same kind. Importantly, image analysis could also aid in identifying other kinds of first- or second-order regulatory components of gene expression, such as cell density or morphology [144, 145]. To integrate new information into the existing mechanistic models, these would need to be updated and modified.

In the studies presented throughout this thesis, we focused on the gene expression regulation of only one or a few genes. However, systems biology today generally tends to a more global view. This change in perspective has been majorly driven by the experimental advancement and development of scRNA-seq [14]. A variety of new single-cell analysis tools for trajectory inference [146–148], pseudo-time estimation [149, 150], and single-cell fate mapping [151] have been established in the recent years, leveraging the rich source of scRNA-seq data. Moreover, algorithms such as SCODE [152] or SCENIC [153] now deploy scRNA-seq data to infer global gene regulatory networks. However, how can scRNA-seq data be linked to mechanism? Gene expression regulation is thought to be driven by a small subset of crucial regulatory genes. While the comprehensiveness of scRNA-seq data enables an unbiased perspective on gene expression regulation, it is currently unknown how to identify the key players important for a mechanistic understanding. A common challenge of big data is finding the correct low-dimensional representation of information. Hence, modern, general machine learning methods such as informed neural networks [154], graph neural networks [155], and autoencoders [156] might help in identifying these key genes. To receive biologically and mechanistically meaningful gene expression kinetics, it is currently also unclear how to map the pseudo-time estimations, i.e, a time estimation inferred by ordering cells according to similar gene expression profiles, underlying scRNA-seq data to real time. Attempts at mapping pseudo-time estimations to real time have been made, however, they are limited to applications with prior knowledge such as cell densities [157]. On the other hand, Ocone et al., fitted mechanistic models on pseudo-time estimations instead of real time [158]. How the gene regulatory networks and reaction rates inferred from scRNA-seq data on a pseudo-time estimation can be translated to real time is unknown and will be dependent on the biological process under consideration. While there are still fundamental questions to be answered before mathematical modeling will be able to leverage all the information contained in scRNA-seq data, to help increase our mechanistic understanding of gene expression regulation, the possibility of it, as such, is exciting.

Finally, only by integrating mechanistic modeling with data from different sources, such as imaging and scRNA-seq, will we be able to fully understand gene expression regulation and its underlying mechanisms.

Bibliography

[1] L. Schuh et al. “Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer”. *Cell Systems* 10.4 (2020), pp. 363–378.

[2] L. Schuh et al. “H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested *Xenopus* embryos”. *Cell Systems* 11.6 (2020), pp. 653–662.

[3] L. Schuh et al. “Gal1 repression memory in budding yeast”. *Submitted* (2022).

[4] C. Alabert et al. “Domain model explains propagation dynamics and stability of histone H3K27 and H3K36 methylation landscapes”. *Cell Reports* 30.4 (2020), pp. 1223–1234.

[5] R. H. Boe et al. “Allelic correlation is a marker of tradeoffs between barriers to transmission of expression variability and signal responsiveness in genetic networks”. *bioRxiv 2021.11.26.470134* (2021).

[6] G. Marion. *Lecture notes: An introductin to mathematical modelling, BioSS*. 2008. URL: https://people.maths.bris.ac.uk/~madjl/course_text.pdf.

[7] G. E. P. Box and N. R. Draper. *Empirical model-building and response surfaces*. John Wiley and Sons, Inc., 1987.

[8] H. Schichl. *Models and the history of modeling*. Springer, Boston, MA, 2004.

[9] J. D. Watson et al. *Molecular biology of the gene*. Benjamin-Cummings, 1965.

[10] J. Zrimec et al. “Learning the regulatory code of gene expression”. *Frontiers in Molecular Biosciences* 8 (2021).

[11] S. Y. Tan and J. K. Furubayashi. “Jacques Lucien Monod (1910-1976): Co-discoverer of the operon system”. *Singapore Medical Journal* 59.10 (2018), pp. 555–556.

[12] G. Orphanides and D. Reinberg. “A unified theory of gene expression”. *Cell* 108.4 (2002), pp. 439–451.

[13] F. Jakob and J. Monod. “Genetic regulatory mechanisms in the synthesis of proteins”. *Journal of Molecular Biology* 3 (1961), pp. 318–356.

[14] F. Tang et al. “mRNA-Seq whole-transcriptome analysis of a single-cell”. *Nature Methods* 6 (2009), pp. 377–382.

[15] A. Raj et al. “Imaging individual mRNA molecules using multiple singly labeled probes”. *Nature Methods* 5 (2008), pp. 877–879.

[16] G. Phillips. “Green fluorescent protein - a bright idea for the study of bacterial protein localization”. *FEMS Microbiology Letters* 204.1 (2001), pp. 9–18.

[17] B. R. Macadangdang et al. “Evolution of histone 2A for chromatin compaction in eukaryotes”. *eLife* (2014).

[18] V. G. Allfrey, R. Faulkner, and A. E. Mirsky. “Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis”. *Proceedings of the National Academy of Sciences of the United States of America* 51.5 (1964), pp. 786–794.

[19] A. J. Bannister and T. Kouzarides. “Regulation of chromatin by histone modifications”. *Cell Research* 21 (2011), pp. 381–395.

[20] T. C. Minshull and M. J. Dickman. “Mass spectrometry analysis of histone post translational modifications”. *Drug Discovery Today: Disease Models* 12 (2014), pp. 41–48.

[21] A. Bastidas-Ponce et al. “Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis”. *Development* 146.12 (2019).

[22] F. Buettner et al. “Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells”. *Nature Biotechnology* 33 (2015), pp. 155–160.

[23] M. Lotfollahi, F. A. Wolf, and F. J. Theis. “scGen predicts single-cell perturbation responses”. *Nature Methods* 16 (2019), pp. 715–721.

[24] A. Ay and D. N. Arnosti. “Mathematical modeling of gene expression: a guide for the perplexed biologist”. *Critical Reviews in Biochemistry and Molecular Biology* 46.2 (2011), pp. 137–151.

[25] B. C. Goodwin. “Oscillatory behavior in enzymmatic control processes”. *Advances in Enzyme Regulation* 3 (1965), pp. 425–438.

[26] J. S. Griffith. “Mathematics of cellular controll processes I. Negative feedback to one gene”. *Journal of Theoretical Biology* 20 (1968), pp. 202–208.

[27] A. Novick and M. Weiner. “Enzyme induction as an all-or-none phenomenon”. *Proceedings of the National Academy of Sciences of the United States of America* 43.7 (1957), pp. 553–566.

[28] A. Raj and A. van Oudenaarden. “Nature, nurture, or chance: stochastic gene expression and its consequences”. *Cell* 135.2 (2008), pp. 216–226.

[29] M. S. H. Ko, H. Nakauchi, and N. Takahashi. “The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates”. *The EMBO Journal* 9.9 (1990), pp. 2835–2842.

[30] M. B. Elowitz, A. J. L. adn Eric D. Siggia, and P. S. Swain. “Stochastic gene expression in a single cell”. *Science* 297.5584 (2002), pp. 1129–1131.

[31] M. S. H. Ko. “A stochastic model for gene induction”. *Journal of Theoretical Biology* 153.2 (1991), pp. 181–194.

[32] A. Raj et al. “Stochastic mRNA synthesis in mammalian cells”. *PLoS Biology* 4.10 (2006).

[33] S. Kesic. “Systems biology, emergence and antireductionism”. *Saudi Journal of Biological Sciences* 23 (2016), pp. 584–591.

[34] S. Aitken, R. D. Alexander, and J. D. Beggs. “Modelling reveals kinetic advantages of co-transcriptional splicing”. *PLoS Computational Biology* 7.10 (2011).

[35] Z. Xu and S. Asakawa. “A model explaining mRNA level fluctuations based on activity demands and RNA age”. *PLoS Computational Biology* 17.7 (2021).

[36] E. C. Dykeman. “A stochastic model for simulating ribosome kinetics in vivo”. *PLoS Computational Biology* 16.2 (2020).

[37] M. A. Ferrin and A. R. Subramaniam. “Kinetic modeling predicts a stimulatory role for ribosome collisions at elongation stall sites in bacteria”. *eLife* 6 (2017).

[38] K. Kozlov et al. “Modeling of gap gene expression in *Drosophila Kruppel* mutants”. *PLoS Computational Biology* 8.8 (2012).

[39] T. R. Sokolowski, T. Erdmann, and P. R. ten Wolde. “Mutual repression enhances the steepness and precision of gene expression boundaries”. *PLoS Computational Biology* 8.8 (2012).

[40] D. M. Holloway et al. “Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in *Drosophila* segmentation”. *PLoS Computational Biology* 7.2 (2011).

[41] R. Giri et al. “Ordered patterning of the sensory system is susceptible to stochastic features of gene expression”. *eLife* 9 (2020).

[42] K. Utsey and J. P. Keener. “A mathematical model of flagellar gene regulation and construction in *Salmonella enterica*”. *PLoS Computational Biology* 16.10 (2020).

[43] C. Zhao et al. “A mechanistic integrative computational model of macrophage polarization: implications in human pathophysiology”. *PLoS Computational Biology* 15.11 (2019).

[44] Y.-Y. Tseng et al. “Comprehensive modelling of the *Neurospora* circadian clock and its temperature compensation”. *PLoS Computational Biology* 8.3 (2012).

[45] S. Magni et al. “Data-driven dynamical model indicates that the heat shock response in *Chlamydomonas reinhardtii* is tailored to handle natural temperature variation”. *Journal of the Royal Society Interface* 15 (2018).

[46] P. L. Freddolino et al. “Stochastic tuning of gene expression enables cellular edaptation in the absence of pre-existing regulatory circuitry”. *eLife* 7 (2018).

[47] S. M. Shaffer et al. “Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance”. *Nature* 546 (June 2017), pp. 431–435.

[48] L. Schuh. “Simulation and characterization of rare correlated gene expression using stochastic network modeling”. MA thesis. Technical University of Munich, 2018.

[49] D. Pokrovsky et al. “A systematic cell cycle block impacts stage-specific histone modification profiles during *Xenopus* embryogenesis”. *PLoS Biology* 19.9 (2021).

[50] N. L. Biggs, E. K. Llyod, and R. J. Wilson. *Graph theory 1736-1936*. Clarendon Press, 1986.

[51] L. Euler. “Solutio problematis ad geometriam situs pertinentis”. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 8 (1736), pp. 128–140.

[52] S. N. Banasode and Y. M. Umathar. “Review of applications of graph theory in engineering”. *International Journal of Mathematics Trends and Technology* 57.4 (2018), pp. 225–230.

[53] T. B. Boffey. *Graph theory in operations research*. Palgrave, London, 1982.

[54] G. A. Pavlopoulos et al. “Using graph theory to analyze biological networks”. *BioData Mining* 4.10 (2011).

[55] F. Riaz and K. M. Ali. "Applications of graph theory in computer science". *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks* (2011), pp. 142–145.

[56] A. Abhishek and S. Singh. "Approaches to gene regulatory network modeling". *International Journal of IT, Engineering and Applied Sciences Research* 2.4 (2013), pp. 35–38.

[57] J. L. Gross, J. Yellen, and M. Anderson. *Graph theory and its applications*. CRC Press LLC, 2018.

[58] J. M. Harris, J. L. Hirst, and M. J. Mossinghoff. *Combinatorics and graph theory*. 2008 Springer Science+Business Media, LLC, 2008.

[59] R. Balakrishnan and K. Ranganathan. *A textbook of graph theory*. Springer New York Heidelberg Dordrecht London, 2012.

[60] R. Aris. "Prolegomena to the rational analysis of systems of chemical reactions". *Archive for Rational Mechanics and Analysis* 19 (1965), pp. 81–99.

[61] P. Érdi and J. Tóth. *Mathematical models of chemical reactions : theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.

[62] U. Alon. *An introduction to systems biology: Design principles of biological circuits*. Chapman, Hall/CRC mathematical, and computational biology series, 2006.

[63] D. J. Wilkinson. *Stochastic modelling for systems biology*. CRC Press, 2011.

[64] P. Mozgunov et al. "A review of the deterministic and diffusion approximations for stochastic chemical reaction networks". *Reaction Kinetics, Mechanisms and Catalysis* 123 (2018), pp. 289–312.

[65] L. Amrhein, K. Harsha, and C. Fuchs. "A mechanistic model for the negative binomial distribution of single-cell mRNA counts". *bioRxiv 657619* (2019).

[66] M. Strasser, F. J. Theis, and C. Marr. "Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression". *Biophysical Journal* 102 (2012), pp. 19–29.

[67] M. K. Strasser et al. "Lineage marker synchrony in hematopoietic genealogies refutes the PU.1/GATA1 toggle switch paradigm". *Nature Communications* 9 (2018).

[68] M. Feinberg. *Foundations of chemical reaction network theory*. Vol. 202. Springer Nature Switzerland AG, 2019.

[69] W. M. Haddad. *A dynamical systems theory of thermodynamics*. Oxford, Princeton University Press, 2019.

[70] J. Hasenauer. "Lecture notes: statistical inference for stochastic biochemical processes, Technical University of Munich". 2017.

[71] C. Guldberg and P. Waage. "Studier i affiniteten". *Forhandelinger: Videnskabs-Selskabet i Christiana* (1864).

[72] D. T. Gillespie. "A rigorous derivation of the chemical master equation". *Physica A* 188 (1992), pp. 404–425.

[73] H. Resat, L. Petzold, and M. F. Pettigrew. “Kinetic modeling of biological systems”. *Methods Mol Biol.* 541 (2009), pp. 311–335.

[74] A. F. Ramos, G. C. P. Innocentini, and J. E. M. Hornos. “Exact time-dependent solutions for a self-regulating gene”. *Physical Review E* 83.6 (2011).

[75] V. Shahrezaei and P. S. Swain. “Analytical distributions for stochastic gene expression”. *Proceedings of the National Academy of Sciences of the United States of America* 11.105 (2008), pp. 17256–17261.

[76] T. G. Kurtz. “The relationship between stochastic and deterministic models for chemical reactions”. *Journal of Chemical Physics* 57 (1972).

[77] D. T. Gillespie. “Approximating the master equation by Fokker–Planck-type equations for single-variable chemical systems”. *Journal of Chemical Physics* 72 (1980).

[78] G. Haag. *Modelling with the master equation: Solution, methods and applications in social and natural sciences*. Springer International Publishing AG, 2017.

[79] B. Munsky and M. Khammash. “A multiple time interval finite state projection algorithm for the solution to the chemical master equation”. *Journal of Computational Physics* 226 (2007), pp. 818–835.

[80] N. G. van Kampen. *Stochastic processes in physics and chemistry*. North Holland; 3rd edition, 2007.

[81] D. T. Gillespie. “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions”. *Journal of Computational Physics* 22 (1976), pp. 403–434.

[82] D. T. Gillespie. “Exact stochastic simulation of coupled chemical reactions”. *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361.

[83] D. T. Gillespie. “Approximate accelerated stochastic simulation of chemically reacting systems”. *Journal of Chemical Physics* 115 (2001).

[84] D. T. Gillespie. “The chemical Langevin equation”. *Journal of Chemical Physics* 113 (2000), pp. 297–306.

[85] D. T. Gillespie and L. R. Petzold. “Improved leap-size selection for accelerated stochastic simulation”. *Journal of Chemical Physics* 119.6 (2003), pp. 8229–8234.

[86] Y. Cao and D. T. Gillespie. “The slow-scale stochastic simulation algorithm”. *Journal of Chemical Physics* 122 (2005).

[87] E. L. Haseltine and J. B. Rawlings. “Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics”. *Journal of Chemical Physics* 117.15 (2002), pp. 6959–6969.

[88] M. A. Gibson and J. Bruck. “Efficient exact stochastic simulation of chemical systems with many species and many channels”. *Journal of Physical Chemistry* 104.9 (2000), pp. 1876–1889.

[89] S. K. Hahl and A. Kremling. “A comparison of deterministic and stochastic modeling approaches for biochemical reaction systems: On Fixed Points, Means, and Modes”. *Frontiers in Genetics* (2016).

[90] D. Schnoerr, G. Sanguinetti, and R. Grima. "Approximation and inference methods for stochastic biochemical kinetics - a tutorial review". *Journal of Physics A: Mathematical and Theoretical* 50 (2017).

[91] A. Raue et al. "Lessons learned from quantitative dynamical modeling in systems biology". *PLoS ONE* 8.9 (2013).

[92] T. G. Kurtz. "Solutions of ordinary differential equations as limits of pure jump Markov processes". *Journal of Applied Probability* 7.1 (1970), pp. 49–58.

[93] W. A. Adkins and M. G. Davidson. *Ordinary differential equations*. Springer Science+Business Media New York, 2012.

[94] K. E. Atkinson, W. Han, and D. Stewart. *Numerical solution of ordinary differential equations*. John Wiley and Sons, Inc., 2009.

[95] S. Engblom. "Computing the moments of high dimensional solutions of the master equation". *Applied Mathematics and Computation* 180.2 (2006), pp. 498–515.

[96] C. H. Lee, K.-H. Kim, and P. Kim. "A moment closure method for stochastic reaction networks". *2009* 130.13 (Journal of Chemical Physics).

[97] J. Hespanha. "Moment closure for biochemical networks". *3rd International Symposium on Communications, Control and Signal Processing* (2008), pp. 42–147.

[98] A. Gandhi, S. Levin, and S. Orszag. "Moment expansions in spatial ecological models and moment closure through Gaussian approximation". *Bulletin of Mathematical Biology* 62.4 (2000), pp. 595–632.

[99] L. Michaelis and M. Menten. "Die Kinetik der Invertinwirkung". *Biochemische Zeitschrift* 49 (1913), pp. 333–369.

[100] G. E. Briggs and J. B. S. Haldane. "A note on the kinetics of enzyme action". *Biochemical Journal* 19.2 (1925), pp. 338–339.

[101] J. Mueller and C. Kuttler. *Methods and models in mathematical biology: deterministic and stochastic approaches*. Springer-Verlag Berlin Heidelberg, 2015.

[102] M. Santillan. "On the use of the Hill functions in mathematical models of gene regulatory networks". *Mathematical Modeling of Natural Phenomena* 3.2 (2008), pp. 85–97.

[103] A. V. Hill. "The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves". *The Journal of Physiology* 40 (1910).

[104] R. H. Singer and D. C. Ward. "Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog". *Proceedings of the National Academy of Sciences of the United States of America* 79 (1982), pp. 7331–7335.

[105] D. J. Wilkinson. "Bayesian methods in bioinformatics and computational systems biology". *Briefings in Bioinformatics* 8.2 (2007), pp. 109–116.

[106] F. Froehlich, C. Loos, and J. Hasenauer. *Scalable inference of ordinary differential equation models of biochemical processes*. Humana Press, New York, NY, 2018.

[107] C. Loos. “Data-driven robust and efficient mathematical modeling of biochemical processes”. PhD thesis. Technical University of Munich, 2019.

[108] L. Held and D. S. Bove. *Applied statistical inference*. Springer-Verlag Berlin Heidelberg, 2014.

[109] H. Hass et al. “Benchmark problems for dynamic modeling of intracellular processes”. *Bioinformatics* 35.17 (2019), pp. 3073–3082.

[110] C. Kreutz. “New concepts for evaluating the performance of computational methods”. *IFAC-PapersOnLine* 49.26 (2016), pp. 63–70.

[111] A. Villaverde et al. “Benchmarking optimization methods for parameter estimation in large kinetic models”. *Bioinformatics* 35.5 (2019), pp. 830–838.

[112] C. Maier, C. Loos, and J. Hasenauer. “Robust parameter estimation for dynamical systems from outlier-corrupted data”. *Bioinformatics* 33.5 (2017), pp. 718–725.

[113] A. C. Hindmarsh et al. “SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers”. *ACM Transactions on Mathematical Software* 31.3 (2005), pp. 363–396.

[114] A. Raue et al. “Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood”. *Bioinformatics* 25.15 (2009), pp. 1923–1929.

[115] O. Chis, J. R. Banga, and E. Balsa-Canto. “GenSSI: a software toolbox for structural identifiability analysis of biological models”. *Bioinformatics* 27.18 (2011), pp. 2610–2611.

[116] A. Villaverde, A. Barreiro, and A. Papachristodoulou. “Structural identifiability of dynamic systems biology models”. *PLoS Computational Biology* 12.10 (2016).

[117] M. P. Saccomani, S. Audoly, and L. D’Angio. “Parameter identifiability of nonlinear systems: the role of initial conditions”. *Automatica* 39 (2003), pp. 619–632.

[118] A. Raue et al. “Comparison of approaches for parameter identifiability analysis of biological systems”. *Bioinformatics* 30.10 (2014), pp. 1440–1448.

[119] P. Stapor et al. “PESTO: Parameter EStimation TOolbox”. *Bioinformatics* 34.4 (2018), pp. 705–707.

[120] K. P. Burnham and D. R. Anderson. “Multimodel inference: Understanding AIC and BIC in model selection”. *Socialogical Methods and Research* 33.2 (2004), pp. 261–304.

[121] A. Blumer et al. “Occam’s razor”. *Information Processing Letters* 24.6 (1987), pp. 377–380.

[122] K. Aho, D. Derryberry, and T. Peterson. “A graphical framework for model selection criteria and significance tests: refutation, confirmation and ecology”. *Methods in Ecology and Evolution* 8 (2017), pp. 47–56.

[123] G. Schwarz. “Estimating the dimension of a model”. *The Annals of Statistics* 6.2 (1978), pp. 461–464.

[124] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer, 2009.

[125] H. Akaike. “A new look at the statistical model identification”. *IEEE Transactions on Automatic Control* 19.6 (1974), pp. 716–723.

[126] C. M. Hurvich. “Regression and time series model selection in small samples”. *Biometrika* 76.2 (1989), pp. 297–307.

[127] S. S. Wilks. “The large-sample distribution of the likelihood ratio for testing composite hypotheses”. *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62.

[128] J. Newport and M. Kirschner. “A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage”. *Cell* 30.3 (1982), pp. 675–686.

[129] G. A. Anderson et al. “Desynchronizing embryonic cell division waves reveals the robustness of *Xenopus laevis* development”. *Cell Reports* 21 (2017), pp. 37–46.

[130] L. Gelens, K. C. Huang, and J. E. F. Jr. “How does the *Xenopus laevis* embryonic cell cycle avoid spatial chaos?” *Cell Reports* 12 (2015), pp. 892–900.

[131] N. Reverón-Gómez et al. “Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication”. *Molecular Cell* 72 (2018), pp. 239–249.

[132] I. Zacharioudakis, T. Gligoris, and D. Tzamarias. “A yeast catabolic enzyme controls transcriptional memory”. *Current Biology* 17 (2007), pp. 2041–2046.

[133] M. Johnston. “A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*”. *Microbiological Reviews* 51 (1987), pp. 458–476.

[134] Y. Li and H. Zhou. “tRNAs as regulators in gene expression”. *Science in China Series C: Life Sciences* 52 (2009), pp. 245–252.

[135] W. Song et al. “Divergent rRNAs as regulators of gene expression at the ribosome level”. *Nature Microbiology* 4 (2019), pp. 515–526.

[136] H. Huang et al. “Lysine benzoylation is a histone mark regulated by SIRT2”. *Nature Communications* 9 (2018).

[137] S. H. Boo and Y. K. Kim. “The emerging role of RNA modifications in the regulation of mRNA stability”. *Experimental and Molecular Medicine* 52 (2020), pp. 400–408.

[138] C. Wang et al. “Real-time imaging of translation on single mRNA transcripts in live cells”. *Cell* 165.4 (2016), pp. 990–1001.

[139] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. *Nature* 521 (2015), pp. 436–444.

[140] C. Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. *Nature Methods* 18 (2021), pp. 100–106.

[141] N. Dietler et al. “A convolutional neural network segments yeast microscopy images with high accuracy”. *Nature Communications* 11 (2020).

[142] F. Padovani et al. “Cell-ACDC: a user-friendly toolset embedding state-of-the-art neural networks for segmentation, tracking and cell cycle annotations of live-cell imaging data”. *bioRxiv 2021.09.28.462199* (2021).

[143] D. J. E. Waibel, S. S. Boushehri, and C. Marr. “InstantDL: an easy-to-use deep learning pipeline for image segmentation and classification”. *BMC Bioinformatics* 22 (2021).

[144] N. Battich, T. Stoeger, and L. Pelkmans. “Control of transcript variability in single mammalian cells”. *Cell* 163 (2015), pp. 1596–1610.

[145] O. Padovan-Merhar et al. “Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms”. *Molecular Cell* 58 (2015), pp. 339–352.

[146] F. A. Wolf et al. “PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells”. *Genome Biology* 20 (2019).

[147] W. Saelens et al. “A comparison of single-cell trajectory inference methods”. *Nature Biotechnology* 37 (2019), pp. 547–554.

[148] Y. Chen, Y. Zhang, and Z. Ouyang. “LISA: Accurate reconstruction of cell trajectory and pseudo-time for massive single cell RNA-seq data”. *Biocomputing* (2018), pp. 338–349.

[149] V. Bergen et al. “Generalizing RNA velocity to transient cell states through dynamical modeling”. *Nature Biotechnology* 38 (2020), pp. 1408–1414.

[150] M. Setty et al. “Characterization of cell fate probabilities in single-cell data with Palantir”. *Nature Biotechnology* 37 (2019), pp. 451–460.

[151] M. Lange et al. “CellRank for directed single-cell fate mapping”. *Nature Methods* (2022).

[152] H. Matsumoto et al. “SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation”. *Bioinformatics* 33.15 (2017), pp. 2314–2321.

[153] S. Aibar et al. “SCENIC: single-cell regulatory network inference and clustering”. *Nature Methods* 14 (2017), pp. 1083–1086.

[154] M. Raissi, P. Perdikaris, and G. E. Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. *Journal of Computational Physics* 378 (2019), pp. 686–707.

[155] F. Scarselli et al. “The graph neural network model”. *IEEE Transactions on Neural Networks* 20.1 (2009), pp. 61–80.

[156] K. D. Yang et al. “Multi-domain translation between single-cell imaging and sequencing data using autoencoders”. *Nature Communications* 12 (2021).

[157] K. Kuritz et al. “Reconstructing temporal and spatial dynamics from single-cell pseudotime using prior knowledge of real scale cell densities”. *Scientific Reports* 10 (2020).

[158] A. Ocone et al. “Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data”. *Bioinformatics* 31 (2015), pp. i89–i96.

A. Core articles

A.1. Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer.

Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer.

Lea Schuh, Michael Saint-Antoine, Eric M. Sanford, Benjamin L. Emert, Abhyudai Singh, Carsten Marr, Arjun Raj, and Yogesh Goyal.

A rare pre-resistant subpopulation of melanoma cells (approximately 1 in 3.000 cells) has been found to survive current targeted drug therapies [47] allowing the surviving cells to aquire resistance and initiate relapse in patients. Shaffer et al. [47] link this rare subpopulation of drug-surviving melanoma cells to extremely high activities (gene expression) in a number of specific marker genes at the initial time of drug therapy. The underlying distributions of these marker gene expressions show long right tails where a pre-resistant cell is found in the long tail of one or several marker gene expression distributions. Additionally, Shaffer et al. [47] have shown that this high gene expression in any of the marker genes is a transient gene expression pattern, arising and disappearing over time.

To identify how these rare transient coordinated high expression states arise in melanoma cells, we described a mathematical framework inlcuding transciptional bursting (the stochastic turning on and off of a gene) and gene interactions to screen 96 different gene regulatory networks (all weakly connected, non-isomorphic and symmetric graphs) containing between 2 and 10 nodes for 1,000 latin hypercube sampled parameter sets resulting in more than 96 million cells simulated with Gillespie's stochastic simulation algorithm. First, we determined well-defined quantitative criteria of the gene expression patterns observed in melanoma and classified each of the resulting simulations into one of four different gene expression classes. Next, we showed that simulations classified as having rare transient coordinated high expression states are qualitatively comparable to experimental data from a drug-naive melanoma population [47]. We then evaluated whether the network topology and/or the parameter set determines the occurrence of simulations with rare transient coordinated high expression states. We found that within a particular network size, the ability to produce rare coordinated high states decreases monotonically with increasing in-degree and auto-regulation. Additionally, we identified a small subset of parameter sets that produced simulations with rare transient coordinated high expression states unproportionally often across all network sizes and topologies. Using a decision tree algorithm and generalized linear models, we found only 3 out of the 7 independent parameters to be critical for producing simulations with rare transient coordinated high expression states. Next, we showed that entry into and exit from the rare transient coordinated high expression states occur through fundamentally different mechanisms, where the marker genes are highly interdependent during the entry but largely independent of each other during the exit. Finally, we hypothesized that resistant cells characterized by a stabilization of high expression may achieve such a stable high expression state by increasing the in-degree of their underlying gene regulatory networks. To validate our hypothesis, we compared the number of edge connections of inferred networks of drug-naive melanoma populations with the number of edge connections of inferred networks of resistant melanoma colonies. Consistent with our hypothesis, inferred networks of drug-naive melanoma populations consisted of substantially more edge connections than the inferred networks of

resistant melanoma colonies.
In this work, we developed a minimal mathematical model of transcriptional bursting and gene interactions which gives rise to gene expression patterns observed in pre-resistant melanoma cells. We showed that the observed gene expression patterns may arise from stochasticity inherent to gene expression rather than highly specialized processes regulating the occurrence of such rare coordinated high states.

Statement of individual contribution

Arjun Raj was interested in investigating the origins of transient gene expression patterns as observed in pre-resistant melanoma cell cultures. Particularly, he asked whether transient gene expression patterns were a result of stochastic gene expression or whether a specific, highly specialized process regulated the occurrence of such rare coordinated high states. In my master thesis, I determined a well defined set of quantitative criteria describing the observed gene expression patterns and formulated an initial mathematical model to produce gene expression patterns as observed in pre-resistant melanoma cell cultures [48]. During my Ph.D., I refined and extended the mathematical model I initially described in my master thesis. Furthermore, I validated the generalizability of the mathematical model with respect to fixed model parameters and the set of quantitative criteria describing the observed gene expression patterns. Using the refined mathematical framework, I generated simulations of gene expression patterns at a large-scale to systemically screen and evaluate gene regulatory networks and parameter sets. I evaluated the resulting simulations and used statistical methods to mechanistically dissect the initiation, maintenance and termination of the observed gene expression patterns. Furthermore, with help from my collaboration partners, we formulated a hypothesis based on the gene expression simulations regarding the formation of stable resistance after drug administration. Our hypothesis was experimentally validated by collaboration partners, where the analysis of chapter *Increasing network connectivity leads to transcriptionally stable states* was performed by Michael Saint-Antoine and Abhyudai Singh. The original manuscript draft excluding the methods section was written by Yogesh Goyal and reviewed as well as edited by me and others. The methods section *STAR METHODS* was written by me and the figures were generated by myself.
I, Lea Schuh, am the main author of this publication.

Permission to include:

Schuh, L., Saint-Antoine, M., Sanford, E.M., Emert, B.L., Singh, A., Marr, C., Raj, A., and Goyal, Y. (2020). Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer. *Cell Systems*, *10*(4), 363-378. doi: https://doi.org/10.1016/j.cels.2020.03.004

The following page contains a copy of the first page of the editorial policies of Cell Systems, including the author's rights. Here is stated, that 'as an author you may [...] include the article in full length or in part in a thesis or dissertation (provided that this is not to be published commercially)' (accessed February 2021). A digital version can be found here: https://www.cell.com/trends/editorial-policies

February 2021, https://www.cell.com/trends/editorial-policies

Editorial Policies

(…)

Authors' Rights

As an author, you (or your employer or institution) may do the following:

- Make copies (print or electronic) of the article for your own personal use, including for your own classroom teaching use.
- Make copies and distribute such copies (including through e-mail) of the article to known research colleagues for the personal use by such colleagues (but not for commercial purposes, as described below).
- Post a revised personal version of the final text (including illustrations and tables) of the article (to reflect changes made in the peer review and editing process) on your personal or institutional website or server, with a link (through the relevant DOI) to the article as published, provided that such postings are not for commercial purposes, as described below. Please note: Depositing in or posting to special repositories (such as PubMed Central or institutional repositories) is permitted only under specific agreements between Elsevier and the repository and only when consistent with Elsevier's policies concerning such repositories.
- Present the article at a meeting or conference and distribute copies of the article to the delegates attending such meeting.
- Allow your employer to use all or part of the information in the article for other intracompany use (e.g., training) if the article is a "work for hire" made within the scope of your employment.
- Retain patent and trademark rights and rights to any process or procedure described in the article.
- Include the article in full or in part in a thesis or dissertation (provided that this is not to be published commercially).
- Use the article or any part thereof in a printed compilation of your works, such as collected writings or lecture notes (subsequent to publication of the article in the journal).
- Prepare other derivative works that extend the article into book-length form or otherwise re-use portions or excerpts in other works, with full acknowledgment of its original publication in the journal.

All copies (print or electronic) or other use of the paper or article must include the appropriate bibliographic citation for the article's publication in the journal. However, you should not indicate in the citation that the version that you are reproducing or posting is the final published version as published in the journal. For example, it may be appropriate to indicate, "This paper has been submitted to [Journal] for consideration."

Commercial purposes include: the posting by companies or their employees for use by customers (e.g., pharmaceutical companies and physician prescribers); commercial exploitation such as associating advertising with such posting (including the linking to advertising by search engines); the charging of fees for document delivery or access; or the systematic distribution to others via e-mail lists or list servers (to parties other than known colleagues), whether for a fee or for free.

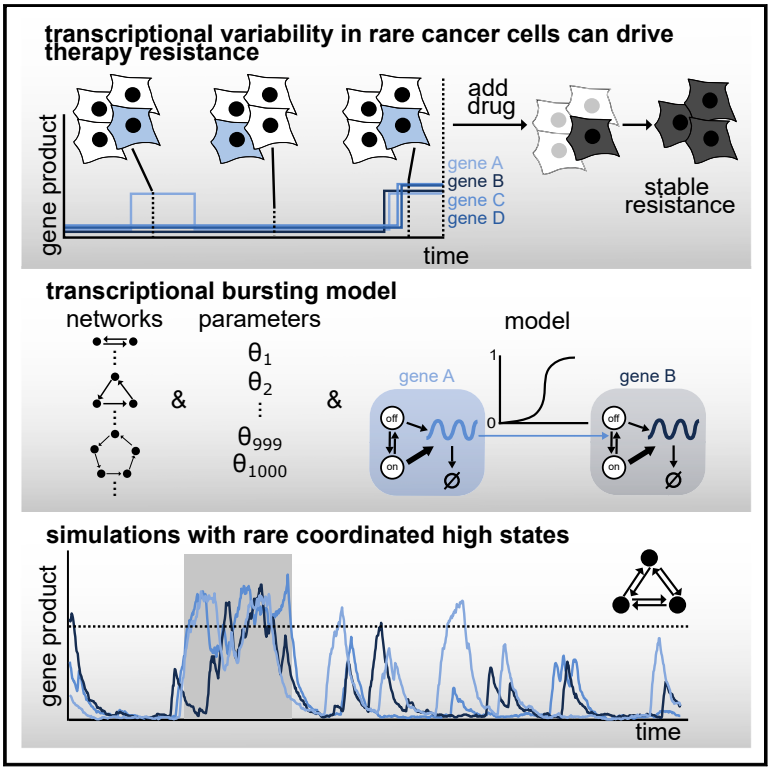
Cell Systems Editorial Policies Version: February 3, 2021

Article

Cell Systems

Gene Networks with Transcriptional Bursting Recapitulate Rare Transient Coordinated High Expression States in Cancer

Graphical Abstract



Authors

Lea Schuh, Michael Saint-Antoine, Eric M. Sanford, ..., Carsten Marr, Arjun Raj, Yogesh Goyal

Correspondence

yogesh.goyal0308@gmail.com

In Brief

Non-genetic transcriptional variability, characterized by transient and coordinated high expression of several genes in rare cancer cells, can drive resistance to targeted therapy. Schuh et al. use a combination of theory and network modeling to demonstrate that established principles of transcription and gene regulation are sufficient to describe the origins of this behavior.

Highlights

- Rare coordinated high expression states in cancer cells can drive therapy resistance
- Gene networks with transcriptional bursting recapitulate these transcriptional states
- Networks with low connectivity favorably give rise to these states
- Parameters affecting transcriptional bursting are critical to produce these states

Schuh et al., 2020, Cell Systems *10*, 363–378
April 22, 2020
https://doi.org/10.1016/j.cels.2020.03.004

CellPress

Cell Systems

CellPress

Article

Gene Networks with Transcriptional Bursting Recapitulate Rare Transient Coordinated High Expression States in Cancer

Lea Schuh,[1,2,3] Michael Saint-Antoine,[4] Eric M. Sanford,[1,6] Benjamin L. Emert,[1,6] Abhyudai Singh,[5] Carsten Marr,[2] Arjun Raj,[1,6] and Yogesh Goyal[1,6,7,*]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA
[2]Helmholtz Zentrum München–German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg 85764, Germany
[3]Department of Mathematics, Technical University of Munich, Garching 85748, Germany
[4]Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19716, USA
[5]Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA
[6]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA
[7]Lead Contact
*Correspondence: yogesh.goyal0308@gmail.com
https://doi.org/10.1016/j.cels.2020.03.004

SUMMARY

Non-genetic transcriptional variability is a potential mechanism for therapy resistance in melanoma. Specifically, rare subpopulations of cells occupy a transient pre-resistant state characterized by coordinated high expression of several genes and survive therapy. How might these rare states arise and disappear within the population? It is unclear whether the canonical models of probabilistic transcriptional pulsing can explain this behavior, or if it requires special, hitherto unidentified mechanisms. We show that a minimal model of transcriptional bursting and gene interactions can give rise to rare coordinated high expression states. These states occur more frequently in networks with low connectivity and depend on three parameters. While entry into these states is initiated by a long transcriptional burst that also triggers entry of other genes, the exit occurs through independent inactivation of individual genes. Together, we demonstrate that established principles of gene regulation are sufficient to describe this behavior and argue for its more general existence. A record of this paper's transparent peer review process is included in the Supplemental Information.

INTRODUCTION

Cellular heterogeneity has been reported to arise from non-genetic transcriptional variability, even in clonal, genetically homogeneous cells grown in identical conditions (Spencer et al., 2009; Sharma et al., 2010, 2018; Gupta et al., 2011; Pisco and Huang, 2015; Fallahi-Sichani et al., 2017; Shaffer et al., 2017; Su et al., 2017). Cells exhibiting these non-genetic deviations are resistant to anti-cancer drugs (e.g., Ras pathway inhibitors) and may lead to relapse in patients. For example, in a drug-naive melanoma population, a small fraction (~1 in 3,000) of cells are pre-resistant, meaning they are able to survive targeted drug therapy, resulting in their uncontrolled cellular proliferation (Shaffer et al., 2017). These rare pre-resistant cells are marked by transient and coordinated high expression of dozens of marker genes. In other words, several genes are highly expressed simultaneously in a rare subset of cells, whereas the rest of the population have low or zero counts of mRNAs for these genes, resulting in a distribution of steady state mRNA counts per cell that peaks at or close to zero and has heavy tails.

The rare cells in the tails, which transiently arise and disappear in the population by switching their gene expression state (Figure 1A), are much more likely to develop resistance to targeted therapies. The rare and coordinated large fluctuations in the expression of multiple genes persist for several generations. Classical probabilistic models of gene expression have predicted the possibility of various types of mRNA expression distributions across a population, including normal, lognormal, gamma, or heavy-tail distributions (Thattai and van Oudenaarden, 2001; Golding et al., 2005; Raj et al., 2006; Raj and van Oudenaarden, 2008; Iyer-Biswas et al., 2009; So et al., 2011; Chen and Larson, 2016; Corrigan et al., 2016; Symmons and Raj, 2016; Antolović et al., 2017; Ham et al., 2019, 2020). It is unclear if such models can recapitulate the non-genetic variability characterized by rare and transient high expression states for several genes simultaneously (from now on referred to as "rare coordinated high states"), and if so, under what conditions.

Might a stochastic system of interacting genes inside the cell facilitate transition in and out of the rare coordinated high state? One hypothesis is that only a rare set of unique (and perhaps complex) networks can facilitate reversible transitions into the rare coordinated high states. Alternatively, relatively generic gene regulatory networks may be capable of producing such behaviors, suggesting that a large

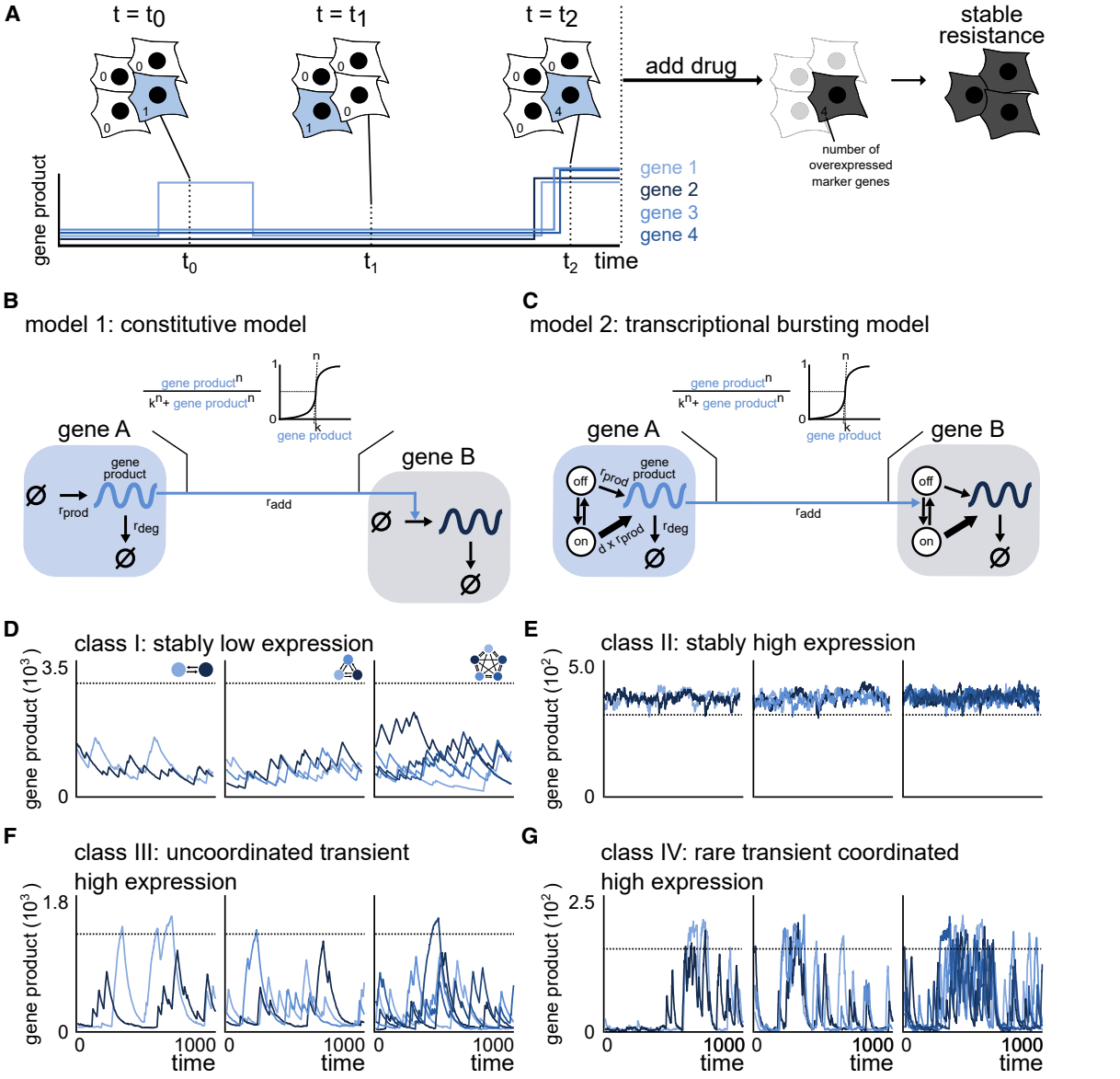


Figure 1. A Transcriptional Bursting Model Is Able to Mimic the Rare Coordinated High States Observed in Drug Naive Melanoma

(A) Drug-naive melanoma cells exist in low (white cells) as well as rare coordinated high (blue cells) expression states. Cells in the rare coordinated high state characterize the pre-resistant state observed in drug-naive melanoma. A schematic of the corresponding expression pattern is shown in the panel below. The cells in a high expression state are more likely to survive and acquire resistance upon drug administration.

(B) Schematic of the constitutive model for two nodes. Gene product is either produced at rate r_{prod} or degraded with rate r_{deg}. Gene regulation is modeled by a Hill function, where the gene product count of the regulating gene A increases the production rate of the gene product of the regulated gene B.

(C) Schematic of the transcriptional bursting model for two nodes. DNA is either in an inactive (off) or active (on) state. Transitions take place with rates r_{on} and r_{off}, where gene product is synthesized with rates r_{pod} and d^*r_{prod}, respectively, $d > 1$. Gene product degrades with rate r_{deg}. Gene regulation is modeled by a Hill function, where the gene expression of the regulating gene A increases the activation of the DNA of the regulated gene B.

(D–G) Depending on the network and the parameters of the transcriptional bursting model, we observe stably low expression (D), stably high expression (E), uncoordinated transient high expression (F), and rare transient coordinated high expression (G). See also Figure S1.

ensemble of such networks may admit rare cell formation. Both of these scenarios have different implications—for instance, the latter hypothesis suggests that this behavior could be more common in biological systems than hitherto appreciated. The alternatives described above can also be posed in terms of the nature of model parameters—whether the set of values that give rise to rare coordinated high states are constrained to lie within a narrow window of parameter space or whether such behavior may occur across broad swaths of parameter space. Yet another possibility is that stochastic gene expression alone fails to produce rare coordinated high states in the absence of additional regulation. In

that case, one may argue that the reversible transition into the rare coordinated high state is driven by highly specialized processes (e.g., initiated by a master regulator) or other unknown mechanisms. Exploring these possibilities will provide potential transcriptional mechanisms that can recapitulate the occurrence of rare coordinated high states.

Here, we describe a mathematical framework to test the hypotheses proposed above for the appearance and disappearance of rare coordinated high states (Box 1). Recent studies from our lab suggest that no particular molecular pathway is solely responsible for the formation of these rare cells (Shaffer et al., 2018; Torre et al., 2019). Specifically, in these rare cells, a sequencing- and imaging-based scheme identified a collection of marker genes, which are targets of multiple signaling pathways ranging from type 1 interferon to PI3K-Akt signaling. The implication is that instead of a single signaling pathway leading to the observed behavior, a network of interacting genes appears to be responsible. Accordingly, we used network modeling to see whether genes interacting within a network were capable of producing transitions to coordinated high expression states. We systematically formulated and simulated networks of increasing size and complexity defined by a broad range for all independent parameters (Boxes 1 and 2; STAR Methods, sections Networks and Parameters).

Computational screens on more than 96 million simulated cells reveal that many networks with interactions between genes are capable of producing rare coordinated high states. Critically, transcriptional bursting, a ubiquitous phenomenon in which genes flip between transcriptionally active and inactive states, is necessary to produce these rare coordinated high states within the context of our models. Subsequent quantitative analysis shows that rare coordinated high states occur across networks of all sizes investigated (up to 10 nodes), but that (1) they depend on three (out of seven) independent model parameters and (2) their frequency of occurrence decreases monotonically with increasing network connectivity. The transition into the rare coordinated high state is initiated by a long transcriptional burst, which, in turn, triggers the entry of subsequent genes into the rare coordinated high state. In contrast, the transition out of the rare coordinated high state is independent of the duration of transcriptional bursts, rather it happens through the independent inactivation of individual genes. We also confirm model predictions using experimental gene expression data (RNA fluorescence in situ hybridization [FISH] data) taken from melanoma cell lines. Together, we demonstrate that the standard model of stochastic gene regulation with transcriptional bursting is capable of producing rare coordinated high states in the absence of additional regulation.

RESULTS

Framework Selection

Identifying the Minimal Network Model Generating Rare Coordinated High States

We focused on a network-based mathematical framework that models cell-intrinsic biochemical interactions and wondered what would be the minimal set of biochemical reactions that constitutes it. Since network models comprised of only constitutively expressed genes were not able to produce rare coordinated high states (Figures 1B, S1A, and S1B; STAR Methods, section Models), we use a leaky telegraph model as the building block of our framework. In terms of chemical reactions, a gene can reversibly switch between an active (r_{on}) and inactive state (r_{off}), where binding of the transcription factor at a gene locus controls the effective rate of gene production (Figure 1C; Box 1; STAR Methods). Specifically, when inactive (or unbound), the gene is transcribed as a Poisson process at a low basal rate (r_{prod}); when active, the rate becomes higher ($d \times r_{prod}$, where $d > 1$). We modeled degradation of the gene product as a Poisson process with degradation rate r_{deg}. The inter-node interaction parameter, r_{add}, has a Hill-function-based dependency on the gene product amount (Hill coefficient n) of the respective regulating node to account for the multistep nature of the interaction (Figure 1C). In particular, we lump steps leading to transcription by implementing the commonly used quasi-equilibrium assumption (Phillips et al., 2019), where binding and unbinding occurs much faster than mRNA transcription and degradation. The dissociation constant k of the Hill function is dependent on the parameters r_{prod}, r_{deg}, and d, such that $k(r_{prod}, r_{deg}, d) = 0.95 \cdot d \cdot \frac{r_{prod}}{r_{deg}}$. In total, the model has seven independent and one dependent model parameters, as outlined in Box 1. All chemical reactions, propensities, and model parameters are presented in STAR Methods. We used Gillespie's stochastic simulation algorithm (Gillespie, 1977) to systematically simulate networks of various sizes and architectures across a broad range of parameters (Box 1; STAR Methods, sections Networks and Parameters).

We limited our study to networks that are symmetric, i.e., networks without a hierarchical structure (Box 1; STAR Methods, section Networks; Figure S1C). We also excluded networks that are compositions of independent subnetworks (non-weakly-connected networks) and networks that can be formed by structure-preserving bijections of other networks (isomorphic networks) (STAR Methods, section Networks; Box 1). These choices reduce the testable space of unique networks by several orders of magnitude (Figure S1C) and allow for comparisons of parameters between networks of different sizes. They also are a conservative starting point for our analysis given experimental observations. In the frequency matrix for experimental RNA FISH data describing the rare high state in drug-naive melanoma, in which each entry corresponds to the fraction of cells with each gene pair being highly expressed (Figure S1D) (Shaffer et al., 2017, 2018), we do not observe a clear directionality of regulation or hierarchical structure within the highly expressed genes. While simulated symmetric networks can recapitulate this experimental observation, asymmetric networks can result in frequency matrices being highly asymmetric (Figures S1E and S1F). For these reasons, we restricted our initial analysis to symmetrical networks.

Characterization of the Transcriptional Bursting Model

When genes are organized in the system described above and simulated over long intervals, the transcriptional bursting model produced a range of temporal profiles for gene products (Figures 1D–1G and S2A). The model was able to faithfully capture the qualitative features of experimental data, i.e., rare, transient, and coordinated high expression states (Figure 1G). We defined a set of rules to screen for the occurrence of different classes of states (Figures 1D–1G and S2A); these include stably low expression (class I), stably high expression (class II), uncoordinated transient high expression (class III), and rare transient coordinated high expression (class IV) (see STAR Methods, section Simulation Classes), and used a heuristic approach to

Box 1. Model Description, Assumptions, Parameters, and Definitions

MODEL DESCRIPTION

The transcriptional bursting model is comprised of single-gene expression modules described by the telegraph model: the DNA can take on an active and inactive state and transcribe mRNA at high and low rates (transcriptional bursting), respectively. These expression modules are coupled by an underlying network architecture, where regulation is modeled by a Hill function: the regulating gene influences the activation rate r_{on} of the respective regulated gene. The chemical reactions and propensities are described below:

Chemical Reaction	Reaction Propensity
$I \rightarrow A$	$\left(r_{on} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}\right) \cdot I$
$A \rightarrow I$	$r_{off} \cdot A$
$I \rightarrow I + mRNA$	$r_{prod} \cdot I$
$A \rightarrow A + mRNA$	$d \cdot r_{prod} \cdot A$
$mRNA \rightarrow \varnothing$	$r_{deg} \cdot mRNA$

where $I,A \in \{0,1\}$, and $I + A = 1$, where $I = 0$ ($A = 1$) denotes that the DNA is in an active state and $I = 1$ ($A = 0$) denotes that the DNA is in an inactive state. $mRNA_X$ is the mRNA count of gene X at the given time. The model aims to recapitulate rare coordinated high states, where "rare" means that at the population level the expression distributions are unimodal and exhibit heavy tails; "coordinated" means that at least once throughout a simulation more than half the genes (nodes) show mRNA expressions above a specified threshold simultaneously; and "high" means that the mRNA expression of a gene exceeds a specified threshold (thres).

MODEL ASSUMPTIONS

(1) mRNA is able to influence the gene expression of its regulated gene directly, hence we refer to it as gene product throughout this work; (2) all genes are relationally identical (weakly-connected, non-isomorphic, and symmetric gene regulatory networks); (3) all genes share the same model parameters; (4) gene regulation is only considered to be activating; and (5) if regulation occurs from several genes, their effects are additive. We discuss and check the generality of our model by testing many of these assumptions on a subset of cases, as described in Box 2.

PARAMETERS

The model is described by 8 model parameters, as defined in the table below along with the corresponding ranges.

Parameters		Sampling Range
Independent Model Parameters		
r_{on}	The rate at which DNA is activated.	0.001–0.1
r_{off}	The rate at which DNA is inactivated.	0.01–0.1
r_{prod}	Synthesis rate of gene product.	0.01–1
r_{deg}	Degradation rate of gene product.	0.001–0.1
r_{add}	Parameter determining the contribution of the additional DNA activation rate upon gene regulation.	0.1–1
d	Factor by which the mRNA synthesis rate is increased when in an active DNA state. $d > 1$.	2–100
n	Hill coefficient.	0.1–10
Dependent Model Parameters		
k*	Dissociation constant of the Hill function, where $k(r_{prod}, r_{deg}, d) = 0.95 \cdot d \cdot \frac{r_{prod}}{r_{deg}}$	-
Dependent Classification Parameters		
thres**	Threshold above which a gene is thought of being highly expressed, where $thres = 0.8 \cdot d \cdot \frac{r_{prod}}{r_{deg}}$	-

(Continued on next page)

Box 1. *Continued*

Here, r_{prod}/r_{deg} is the steady state in the baseline expression state (when there is no transcriptional burst), and d^*r_{prod}/r_{deg} is the steady state in the high expression state (if the DNA would continuously be in the active state).

MODEL DEFINITIONS

- Weakly connected network: a directed network that when replacing the directed edges by undirected ones produces a connected graph in which every pair of nodes is connected by a path.
- Non-isomorphic: two graphs are called non-isomorphic if there exists no structure-preserving bijection between them.
- Symmetric: within a graph the number of in- and outgoing edges of a node and across nodes is identical and either all nodes in a network have a self-loop or not.
- Rare coordinated high state: (1) at least once within a simulation more than half the genes are highly expressed simultaneously, (2) the histogram of simultaneously highly expressed genes at the population level decreases, and (3) the gene expression distributions at the population level are heavy-tailed.
- Connectivity: number of ingoing edges for any node of the network.
- Characteristic distance: the average shortest path length between pairs of nodes of the network.

*The parameter k is dependent on the parameters r_{prod}, r_{deg}, and d, such that: $k(r_{prod}, r_{deg}, d) = x \cdot d \cdot \frac{r_{prod}}{r_{deg}}$, where $x \in \{0.75, 0.8, 0.85, 0.9, 0.95, 1\}$, which ensures a consistent definition of k throughout the network architectures and parameter sets. Here, x represents the fraction of the value corresponding to the steady state value in the high expression state. We showed that for x = 0.75, none of the 100 simulations show rare coordinated gene expression because the threshold resulting in an effective gene regulation is exceeded too often—the regulated DNA states are activated more frequently leading to the high gene expression states and loss of rareness of the coordinated high gene expression event (leading to bimodal distributions). For $x > 0.75$, there is an increase in the number of simulations showing rare behavior, peaking at x = 0.95. Furthermore, throughout different values of x, the same parameter sets give rise to rare coordinated high states. We take x = 0.95 to maximize the number of simulations positive for the rare coordinated high states.

**We test several values for the threshold above which a gene is highly expressed: $thres = y \cdot d \cdot \frac{r_{prod}}{r_{deg}}$, where $y \in \{0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\}$. For all $y \geq 0.6$, the set of simulations showing rare coordinated high states largely remains the same. Even for y = 0.3, half of the simulations identified previously to show rare behavior are still classified as such. We chose x = 0.8. Though arbitrarily chosen, the choice of x = 0.8 will not change the conclusions of our analysis.

distinguish between these different classes (Boxes 1 and 2). For a detailed description of the rules and quantitative metrics used to define class IV, see Boxes 1 and 2; Figures S3 and S4; STAR Methods, section Simulation Classes.

To better compare the computational results with the experimental data from static RNA FISH images, we split the entire simulation into non-overlapping time intervals of 1,000 time units, as justified by the ergodic theory (Box 2; STAR Methods) (Van Kampen, 1992). We took snapshots of gene products at randomly selected time-points in these time intervals and noted the number of simultaneously highly expressed genes as well as their gene product counts, allowing us to represent the static states of a population of simulated cells (Figure 2A). For example, in a particular 8-node network, we found that the distribution qualitatively captures the experimental observations where most cells do not exhibit high expression states, whereas some cells are in a high state for one or more genes (Figure 2B). Similarly, when we selected a gene and plotted its product count for the randomly selected time-points, we observed a heavy-tailed distribution (Figure 2C, right panel), similar to the experimental observations (Figure 2C, left panel). These observations, while shown for a particular 8-node network, also hold true for simulations of other 8-node networks as well as networks of other sizes (Figure S2B).

Note that the simulated distributions of gene product counts for each gene are qualitatively similar because each gene is equivalent within our symmetrical networks (Figure S2C). This is not biologically realistic; the experimental data in drug-naive melanoma cells for mRNA counts display different degrees of skewness of the distribution for different genes (e.g., epidermal growth factor receptor [EGFR] versus Jun, Figure S3A) (Shaffer et al., 2017). These experimental observations can be recapitulated in the simulated networks by introducing asymmetries. For example, two asymmetric networks we tested were able to produce rare coordinated high states (Figures S2G–S4M) and distributions of gene product counts with different degrees of skewness (Figure S2M). When experimentally observed expression distributions (Figure S3A) are compared to simulated expression distributions using Gini coefficients, we observe that while the Gini coefficient is low for most of the simulations (99.2%, gray), it is much higher for the simulations that produce rare coordinated high states (red) and overlaps with experimental Gini coefficients observed for individual genes (Figure 2D). In total, these observations suggest that a simple transcriptional bursting model is able to produce states that recapitulate key aspects of rare coordinated high states observed in drug-naive melanoma.

Rare Coordinated High States Depend on Network Topologies and Model Parameters

Since the rare coordinated high states occur in <1% of all simulations (Figure S2A), we wondered whether their occurrence depends on the network topologies and/or model parameters.

Box 2. Relaxing Model Assumptions

PROTEIN TRANSLATION

The original transcriptional bursting model does not include a step for translation and is assumed to be captured by the Hill function term which not only greatly reduces the computational costs of long stochastic simulations but also allows for analyzing smaller sets of parameters. To check if our model can produce rare coordinate high states even when the model includes the translation step, we focused on a particular network (5.3) and associated parameter values that give rise to these states in the original model. We show that for specific rates of translation and protein degradation (STAR Methods), the model including translation exhibits the rare coordinated high states.

NETWORK ARCHITECTURES

By reducing the network architectures to weakly connected, non-isomorphic, and symmetric networks, we systematically reduce the number of possible network architectures. The reduced space of networks is partly supported by experimental observations (Shaffer et al., 2017, 2018), reporting that (1) there is no obvious hierarchical relationship between the expressed genes; and (2) no particular signaling pathway appears to be solely responsible for the observed behavior (see also Figure S1D). Furthermore, these network architectures allow for direct comparisons between network sizes, connectivities, and parameter sets (not a given for other topologies). Although the analysis here primarily focuses on the constrained set of network architectures, we show for a subset of cases (STAR Methods) that asymmetric network architectures can also exhibit rare coordinated high gene expression states (Figures S2G–S2I), paving the way for a more systematic analysis in the future studies.

MODEL PARAMETERS

While we primarily focus on keeping the same parameter set for each node, we analyzed a subset of networks with asymmetric parameters (STAR Methods) such that each node had distinct underlying parameter sets. We show that a model with asymmetric parameter sets is also capable of producing rare coordinated high gene expression states (Figures S2J–S2M).

MULTI-GENE REGULATORY EFFECTS

The joint regulatory effects experienced by a gene, which is regulated by several other genes, can be modeled using different approaches. While the majority of analysis here uses an additive model of joint-regulation, we performed a subset of simulations (STAR Methods) for cases where the regulation by multiple gene nodes is multiplicative (Figures S4C and S4E). We find that for network architecture 5.3, 15 and 97 out of 1,000 parameter sets give rise to simulations with rare coordinated high states in the additive and multiplicative joint-regulation, respectively (Figure S4D). Nine simulations are found to show rare coordinated high states in both definitions of multi-gene regulation.

DEFINING MODEL-OUTPUT METRICS

Population Level—Sub-simulation Size to Determine a Single Cell

To qualitatively compare our results to experimental data, we convert the 1,000,000 time units long single-cell simulation to 1,000 single-cell sub-simulations of length 1,000 time units. We show that the simulations are largely (88.2%) uncorrelated after 1,000 time units, justifying our analysis (STAR Methods).

Heavy-tails

We test different levels of stringency in our definition of heavy-tailed or sub-exponential distributions. The analysis in Figures 2 and 3 is performed using the criteria described in STAR Methods, section Simulation Classes. We perform further analysis similar to Figures 2 and 3 by using more stringent definitions, i.e., fit exponentials and compare the 99th percentiles (Figure S3C). We demonstrate that these results and conclusions are similar to the ones obtained using less stringent criteria (Box 1) shown in Figures 2 and 3 (see Figures S4F–S4M). For example, 6 and 7 out of 8 rare coordinated high parameter sets also appear in the two more stringent analyses (Figures S4H and S4L). We further validate that our model recapitulates the experimentally observed heavy-tails by comparing the Gini coefficients (Jiang et al., 2016) of experimental and model distributions (Figure 2D).

NUMBER OF NODES HIGHLY EXPRESSED TO BE CALLED A "COORDINATED" STATE

We define a simulation to show coordinated high gene expression if at least once throughout the simulation more than half of the gene product counts exceed the threshold. Furthermore, we show that for different node counts (2, 3, 4, 5) the number of simulations showing rare coordinated high states does not vary significantly. As an example, for a count of 2, we get 6 out of 100

(Continued on next page)

Box 2. *Continued*

simulations showing rare behavior; for a count of 3, we get 7. Note that the sets of simulations were overlapping between different scenarios.

DEFINITION OF RARE COORDINATED HIGH PARAMETER SETS

We define rare coordinated high parameter sets as parameter sets showing rare coordinated high expression in $\geq$ 20% of all 96 networks. The threshold was defined by inspecting the histogram (Figure 3A), where we see a separation at 20%. Notably, the same rare coordinated high parameter sets also appear in other analyses—they show increased frequencies of simulations with rare coordinated high states when considering the network sizes separately (Figure S6A). Additionally, stricter definitions for heavy-tailed expression distributions result in similar rare coordinated high parameter sets (Figures S4H and S4L).

BOOTSTRAPPING CONTROLS IN PHIXER ALGORITHM

As the number of connections predicted by the Phixer algorithm can depend on the sample size, we bootstrapped the original data set into 4,000-sample datasets. The number 4,000 was chosen arbitrarily; bootstrapped sample sizes of 1,000, 2,000, and 6,000 also produced qualitatively similar results.

EDGE WEIGHT IN PHIXER ALGORITHM

We created a randomized control consisting of permutations of each gene column from the original dataset. We then performed the Phixer analysis on these randomized controls. The resulting edge weight distributions give us a baseline or control edge weight for Phixer that, in principle, reflects potential false positives. We found that in the controls, nearly all of the predicted edge weights were below 0.45 (Figure S8B). Therefore, we decided to choose 0.45 as a threshold for our non-control analysis, thus eliminating edges that could have been predicted by chance alone.

Specifically, what are the features of the topologies and parameters that facilitate the occurrence of rare coordinated high states? For the simulations that produced rare coordinated high states, we extracted and quantitatively analyzed the corresponding networks. We found that the rare coordinated high states occur ubiquitously in networks with different numbers of nodes analyzed (up to 10 nodes) (Figures 2E, S2B–S2F, S5A, and S5B). Within a particular network size, the ability to produce rare coordinated high states decreases monotonically with increasing network connectivity (Figures 2F, S5C, and S5D). Consistently, the fraction of networks per network size (normalized by either network size or total networks per network size) exhibiting rare coordinated high states decreases with increasing size (Figures S5A and S5B) as a larger fraction of high connectivity networks exist in bigger networks (Figure S5D).

We next wondered whether gene auto-activation (networks with self-loops) have any effect on a networks ability to produce the rare coordinated high states. We found that adding self-loops on otherwise identical networks reduced the occurrence number of simulations with rare coordinated high states (Figure 2G). We also analyzed network topologies based on characteristic distance, defined as the average shortest path length between pairs of nodes of the network (see STAR Methods; Box 1). Characteristic distance recapitulates the effects of not only network connectivity (inversely correlated with characteristic distance) but also differentiates topologies with the same connectivity (Figure 2H), for example, networks with or without auto-activation. Using this metric across networks of all sizes, we found that higher numbers of simulations exhibit rare coordinated high states for larger characteristic distances. Together, we demonstrate that the occurrence of rare coordinated high states depends on network topologies.

Since the transcriptional bursting model has seven independent parameters (r_{on}, r_{off}, r_{prod}, r_{deg}, r_{add}, d, and n; see Box 1 for details), we asked whether specific parameter combinations preferentially give rise to the rare coordinated high states, and if so, what features of such combinations facilitate it. The subsequent analysis is motivated by the initial observation that occurrence of different classes of temporal gene product profiles across different network sizes and connectivities appear to also depend on the parameter sets (Figure 2I). Specifically, if a parameter set gave a specific expression profile (e.g., rare coordinated high or stably high) for one network, it displayed a higher propensity to display the same profile for other networks as well (Figures 2I and S3D), implying that parameters indeed play a major role in the occurrence of rare coordinated high states. To avoid biases in the parameter sets investigated, all 1,000 parameter sets were sampled from a broad range for each parameter using a latin hypercube sampling algorithm (Table S1; STAR Methods, section Parameters).

We first measured the percentage of simulations per parameter set that gave rise to the rare coordinated high states. Out of the 1,000 parameter sets, eight parameter sets, from now on called rare coordinated high parameter sets (Box 2), clustered together at the tail end of the distribution (orange, Figure 3A), meaning they generated simulations with frequent occurrence of rare coordinated high states in at least 20% of all networks tested (Figure 3A). Furthermore, these eight parameter sets robustly generated rare coordinated high states across all network sizes and architectures (Figure S6A). Therefore, we wondered if these eight parameter sets have any special or

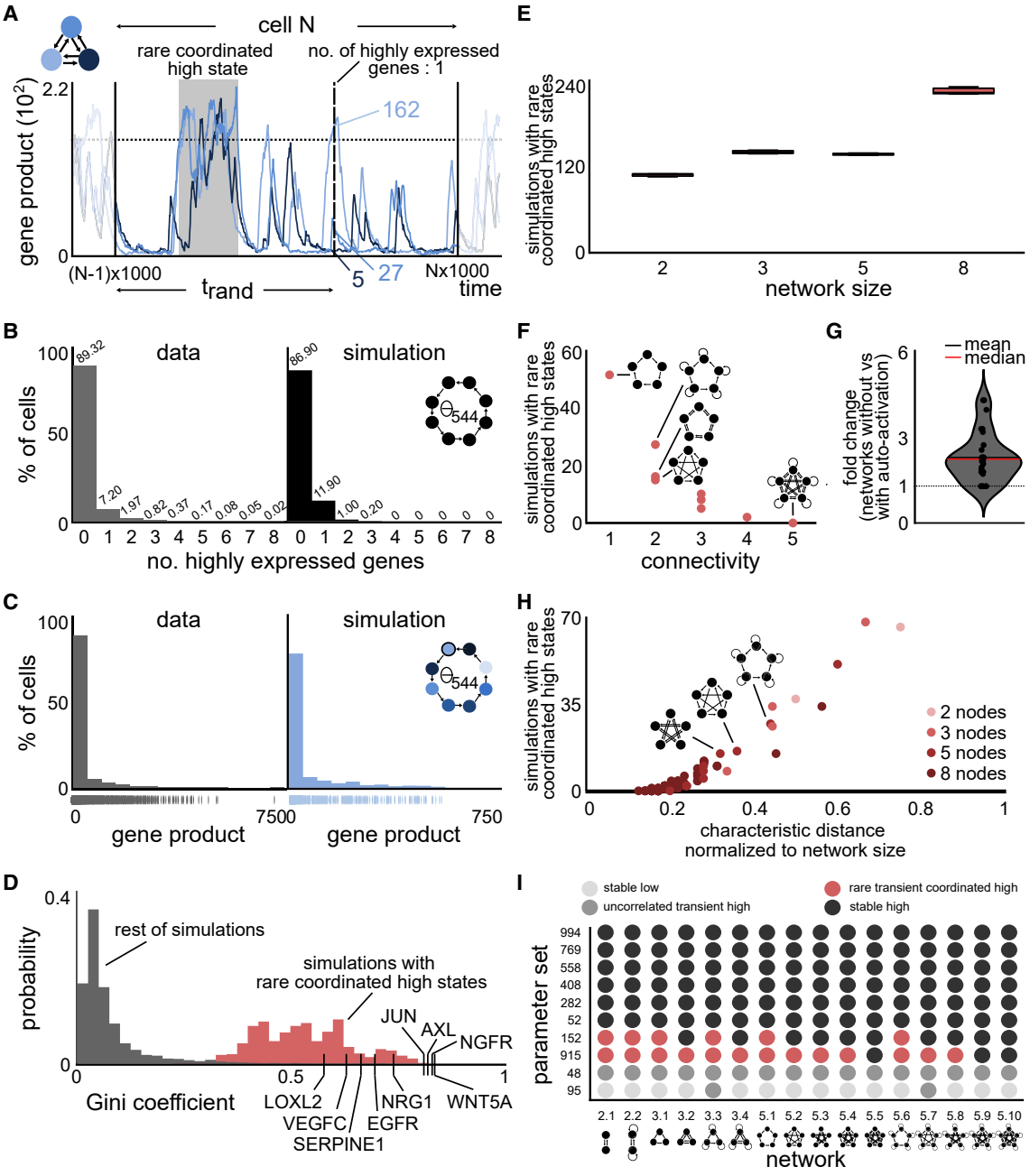


Figure 2. Simulations of the Transcriptional Bursting Model Show Similar Behavior at the Population Level as the Drug-Naive Melanoma Cells

(A) Frame of a simulation showing rare coordinated high state (shaded area). The 1,000,000 time unit simulation is split into frames of 1,000 time units to create a simulated cell population (shown for cell N). For a randomly determined time point t_{rand}, the number of simultaneously highly expressed genes and the gene count per gene per cell are evaluated. The network of the corresponding simulation is given in the top left corner.

(B and C) The simulated number of simultaneously highly expressed genes and expression distribution at the population level are qualitatively similar to experimental data from a drug-naive melanoma population (data from Shaffer et al., 2017). The percentages are indicated above the histogram (in B). The network and parameter set as well as the particular node (in C) used for comparison are shown in the right panel.

(D) The Gini indices of simulations of rare coordinated high states are substantially higher than of simulations not showing rare coordinated high states. The experimentally measured expression distributions have Gini indices similar to simulations with rare coordinated high states.

(E) Total number of rare coordinated high states were extracted for simulations of different networks sizes, containing either 2, 3, 5, or 8 nodes to see if they occur across networks of different sizes. Rare coordinated high states were found to exist ubiquitously across all possible networks of all analyzed network sizes. The measurements were performed via three independent and randomly sampled t_{rand} (median, 25th and 75th percentiles).

(legend continued on next page)

distinguishing features compared to the remaining 992 parameter sets.

We used a decision tree algorithm (Breiman et al., 1984) (see STAR Methods, section Decision Tree Optimization and Generalized Linear Models) to identify the differentiating features of the rare coordinated high parameter sets from the rest. The decision tree analysis revealed that only three (r_{on}, r_{off}, and r_{add}) of the seven independent parameters showed a strong correlation with the rare coordinated high parameter sets (Figure 3B). We validated these findings with complementary analysis using generalized linear models (STAR Methods, section Decision Tree Optimization and Generalized Linear Models) where we found precisely these three specific parameters (r_{on}, r_{off}, and r_{add}) to be critical to produce the rare coordinated high states with high statistical significance (p values: r_{on} = 0.003; r_{off} = 0.005; r_{add} = 0.014) (Figure S6B). These observations became readily evident when we plotted all 1,000 parameter sets for r_{on}, r_{off}, and r_{add} together and found that the rare coordinated high parameters sets occupy a narrow region of the parameter phase space (Figures 3C and S6C). These three parameters are related to transcriptional bursting and inter-gene(node) regulation. Two of these parameters, r_{on} and r_{off}, define the transitioning between the active and inactive state of the DNA respectively. The third parameter is the gene activation rate, r_{add}, which corresponds to the positive regulation of transcriptional bursting rate of a gene by the gene product of another interacting gene. Parameter sensitivity analysis across the parameter space also confirmed that these three parameters are indeed critical for producing the rare coordinated high states (Figure S6D). Too high values (> 0.31) of r_{add} result in the disappearance of rare coordinated high states, as does a complete absence (r_{add} = 0) of this term (Figures S6E–S6G). To confirm that these three parameters (r_{on}, r_{off}, and r_{add}) and their corresponding range of values are indeed critical to producing simulations with rare coordinated high states, we sampled 1,000 new parameter sets from a constrained region containing all eight rare coordinated high parameter sets (Figure 3C, orange box; STAR Methods) and ran simulations for two test networks, a 3-node and a 5-node network. We found that the frequency of simulations with rare coordinated high states for the constrained region is ~14-fold and ~21-fold higher than that for the original parameter space, respectively (Figure 3D). We note that although parameter sets with parameters r_{on}, r_{off}, and r_{add} within the identified critical parameter ranges give rise to simulations with rare coordinated high states much more frequently than other parameter sets, it is not 100% of the time.

Distinct Mechanisms Regulate the Transition into and out of Rare Coordinated High States

We have identified the networks and parameter sets for which the transcriptional bursting model exhibits rare coordinated high states more frequently. Next, we dissected the features of the model that facilitate the occurrence of rare coordinated high states. Specifically, we identified the factors that (1) trigger the entry into the rare coordinated high states, (2) facilitate its maintenance, and (3) trigger the escape from it. We began by analyzing various features of transcriptional activity, since including transcriptional bursting was found to be critical for the model to display the rare coordinated high states. These include the burst fraction, length of transcriptional bursts (burst duration), and burst frequency. To measure these features, we defined four regions for each simulation: low expression state (baseline time-region), entry into the high expression state (entry time-point), the high expression state (high time-region), and exit from the high expression state (exit time-region) (Figure 4A; STAR Methods, section Entry and Exit Mechanisms).

We found an increase in the transcriptional activity, as measured by the burst fraction, during the high expression time-region compared with the baseline time-region (Figure 4B), suggesting that enhanced transcriptional activity facilitates the maintenance of rare coordinated high states. Increased burst fraction could be a result of (1) longer transcriptional bursts or (2) a higher burst frequency. The former is not possible as the duration of each burst is distributed exponentially according to exp(r_{off}), which does not change between the baseline and high time-region. Indeed, we found an increase in the burst frequency in the high time-region, thus establishing its role in the maintenance of the rare coordinated high state (Figure 4C). The increased transcriptional bursting seen in the models capable of generating rare coordinated high states is consistent with the experimental observations that the transcriptional activity occurred in frequent bursts in cells high for a marker gene (Shaffer et al., 2018). Next, we wondered whether burst frequency increases with the interactions of genes within the network. We compared two networks of the same size (3 nodes), where one is comprised out of single unconnected (orphan) nodes and the

(F) The frequency of rare coordinated high states depends on the network connectivity, which is defined as number of ingoing edges for any node of the network. Shown here is the dependence for all 5-node networks, such that increasing connectivity within all 5-node networks leads to a decrease in the number of simulations with rare coordinated high states. Each dot represents a particular network topology within the possible space of 5-node networks.

(G) Effect of adding auto-activation (self-loop) to networks on the number of simulations with rare coordinated high states. Networks with auto-activation exhibit simulations with rare coordinated high states less frequently than the same networks without auto-activation. Fold change is calculated by dividing the number of simulations with rare coordinated high states for networks containing auto-activation with the number of simulations with rare coordinated high states for the same networks without auto-activation. Each dot represents one of the 96/2 = 46 direct network comparisons. Network comparisons where one of the networks did not give rise to simulations with rare coordinated high states were discarded.

(H) The frequency of simulations with rare coordinated high states depends on the characteristic distance, defined as the average shortest path length between pairs of nodes of the network. With increasing characteristic distance (normalized to network size), more simulations show rare coordinated high states. Each dot represents the characteristic distance of one of the 96 networks. Each network size is represented by a unique color.

(I) The frequency of occurrence of simulations with rare coordinated high states is dependent on the choice of model parameters. Specifically, simulations of a particular parameter set across different networks and sizes show largely the same class of gene expression profiles. Each row corresponds to specific parameter sets within the space of all parameter sets analyzed. Each column name corresponds to a particular network, and the underlying network is drawn below the column name.

See also Figures S2–S5.

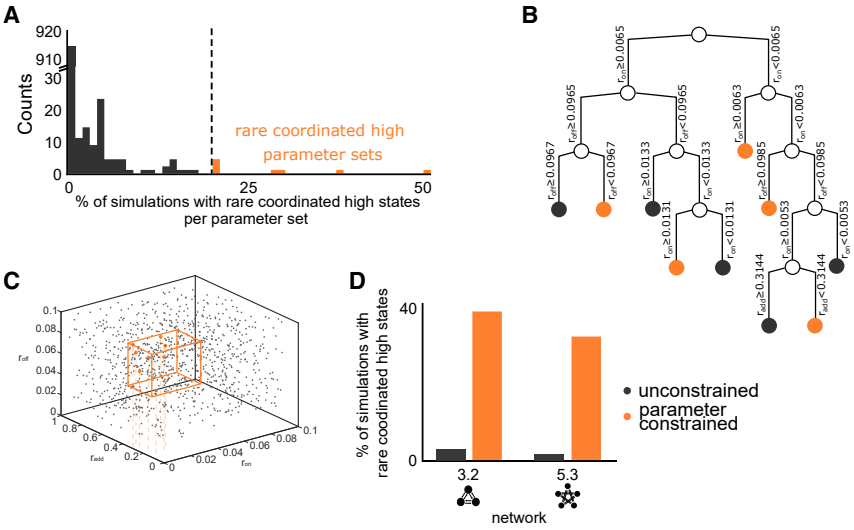


Figure 3. Transcriptional Bursting Rates Influence the Formation of Rare Coordinated High States

(A) Histogram of the percentage of simulations with rare coordinated high states per parameter set to identify the parameter sets that favorably give rise to simulations with rare coordinated high states. Each of the 96 networks is simulated for every single of the 1,000 parameter sets, where not all 96 of these simulations give rise to rare coordinated high states. The eight rare coordinated high parameter sets, marked in orange, produce rare coordinated high states in more than 20% (more than 19 out of the 96 simulations) of simulations and lie at the tail of the histogram. The cutoff (dashed line) marks the 20%.

(B) Decision tree optimization was performed to identify differentiating features of the rare coordinated high parameter sets (orange in Figure 3A) from the rest (dark gray in Figure 3A). Decision tree analysis revealed that only three out of seven parameters, r_{on}, r_{off}, and r_{add}, show a strong correlation with the rare coordinated high parameter sets. Each arm represents a decision, where the decision is marked on top, and each colored dot represents a final class.

(C) Three-dimensional representation of all tested 1,000 parameter sets for r_{on}, r_{off}, and r_{add} show that the rare coordinated high parameter sets are narrowly constrained in the 3D space (orange dots). The orange box indicates the constrained parameter space enclosing all rare coordinated high parameter sets used for analysis in (D).

(D) Comparison between the original 1,000 parameter sets and 1,000 parameter new sets sampled from the constrained region (orange box in Figure 3C) containing all eight rare coordinated high parameter sets. As compared to the original parameter sets, constrained region parameter sets strongly favor the formation of rare coordinated high states for both of the networks tested (3.2 and 5.3). 3.2 and 5.3 correspond to particular networks (outlined below each bar) of network size three and five, respectively.

See also Figures S4 and S6.

other of an interdependent structure (network 3.2). We found that for any parameter set (screened for all 26 parameter sets giving simulations with rare coordinated high states in the previous analysis for network 3.2, Table S1), the system with a connected network has (1) more high expression states and (2) prolonged time in high expression states, as compared with unconnected nodes (Figure 4D). Together, we find that the maintenance of the high state is because of increased burst frequency.

Next, we wanted to identify the factors triggering the entry into the rare coordinated high states. We found that for any gene in the network, the transcriptional burst duration right before and/or during the entry into a rare coordinated high state was significantly higher (two-sample Kolmogorov-Smirnov test) than that in the baseline time-region (i.e., regular bursting kinetics). In the example shown in Figure 4E, the average time of transcriptional burst at the entry time-point is 84.82 (time units) as compared with only 15.08 (time units) in the baseline time-region. Therefore, prolonged transcriptional bursts play a role in driving the cell to a coordinated high expression state. Conversely, we asked if the opposite is true at the exit time-region, such that transcriptional bursts for the exit time-region are shorter than for the high time-region. We found no statistical difference in the distributions of burst durations between the high and the exit time-regions, as demonstrated by the example in Figure 4F, suggesting that the exit from high expression state occurs independently of the burst durations. Both of these conclusions hold true when measured for all simulations with rare coordinated high states (Figure 4G). Together, unlike the entry into the high time-region, the exit from it is not dependent on the transcriptional burst duration.

We also wondered if the entry into the high expression state of one gene influences the entry of other genes, or that the genes enter the high expression state independently of each other. We reasoned that if the time duration between two successive genes (t_{ent}, Figure 4A) entering the high expression state is exponentially distributed, it would imply that the genes enter the high expression state independent of each other. Instead, we found that the distributions of entry time intervals rejected the null hypothesis of the Lilliefors' test for most of the simulations (84%), meaning they are not exponentially distributed (Figure 4H). The remaining 16% of cases were found to be largely falsely identified as exponentially distributed because of limited data (see a representative example in Figure S7A). Similarly, we tested if the exit for successive genes from the high expression state occurs independent of each other. Contrary to the situation during the entry into the high expression state, many distributions of exit time intervals satisfied the null hypothesis of the Lilliefors' test, implying they are indistinguishable from exponential distributions (Figure 4I). The simulations that did not satisfy the stringent Lilliefors' test mainly appear to be exponentially distributed; nevertheless, a representative example is shown in Figure S7B. Together, the entry into and exit from the rare coordinated high state occur through fundamentally different mechanisms—the entry of one gene into the high expression state affects entry of the next gene, while they exit from it largely independently of each other. The exit from the high state could be a result of weak strength of coupling (as reflected by the moderate values of parameter r_{add}) between nodes for the simulations that produce these states. Consistently, we found that too high values of r_{add} result in the disappearance of rare coordinated high

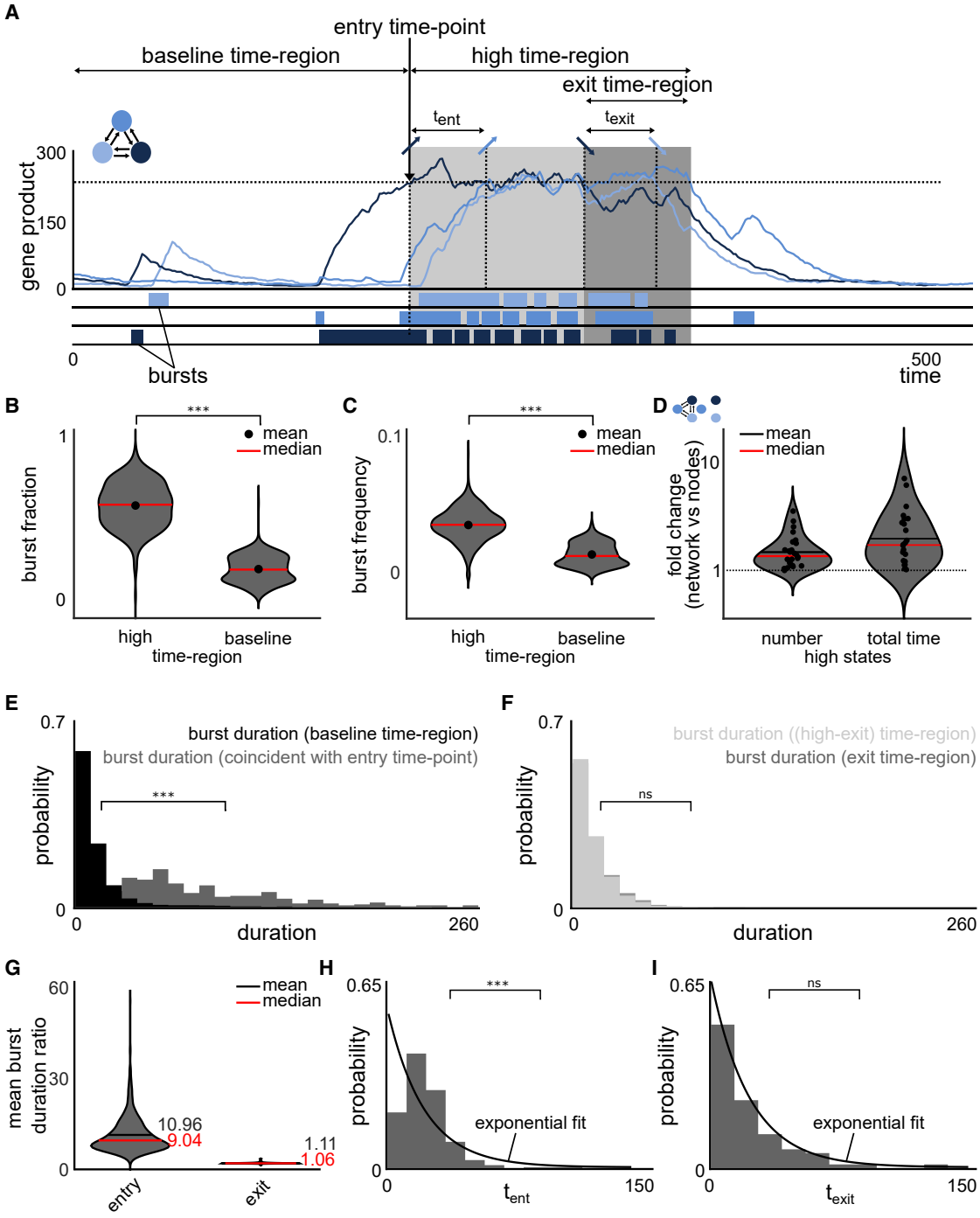


Figure 4. Rare Coordinated High State is Initiated by a Long Transcriptional Burst, Maintained by an Increase in Burst Frequency and Terminated According to a Random Process

(A) An exemplary high region, with a baseline time-region, entry time-point, high time-region, and an exit time-region. The time intervals for an additional gene to enter and exit the high region are marked by t_{ent} and t_{exit}, respectively. The bursts below the exemplary simulation are representative schematics.

(B) Burst fraction, defined as the number of time-points the system is in a burst divided by the total number of time-points, was calculated for baseline time-region and high time-region for all (n = 594) simulations that produce rare coordinated high states and compared using violin plots. The burst fraction is significantly higher in the high time-region than the baseline time-region (two-sample Kolmogorov-Smirnov test, p value < 0.001), implying that enhanced transcriptional activity facilitates the maintenance of rare coordinated high states.

(C) Burst frequency, defined as the number of bursts divided by the total number of time-points, was calculated for baseline time-region and high time-region for all (n = 594) simulations that produce rare coordinated high states and compared using violin plots. The frequency of transcriptional bursts is increased in the high

(legend continued on next page)

states, giving way to stable high states. In other words, the network can transition into the high expression state but loses the ability to come out of it (Figures S6E–S6G).

Increasing Network Connectivity Leads to Transcriptionally Stable States

So far, we have used the transcriptional bursting model to understand the potential origins of rare pre-resistant states in drug-naive melanoma cells. Upon treatment with anti-cancer drugs, the transient pre-resistant cells reprogram and acquire resistance resulting in their uncontrolled proliferation. The resistant cells are characterized by the stabilization of the high expression of the marker genes, which were transiently high in the drug-naive pre-resistant cells (Figure 5A) (Shaffer et al., 2017). Studies using network inference of gene expression data have suggested that the genetic networks undergo considerable rearrangements upon cellular transitions or reprogramming (Moignard et al., 2015; Schlauch et al., 2017). We wondered if the transcriptional bursting model can explain how the transient high expression in drug-naive cells might become permanent upon treatment with anti-cancer drugs. The modeling framework produces a range of gene expression profiles, depending on the network properties and model parameters (Figures 1D–1G). Increasing the network connectivity (for fixed parameter sets) is one way to shift from a rare transient coordinated high expression state to stably high expression state (Figures 5B–5E). As an example, for a fixed network size (five) and associated parameters, increasing the network connectivity from one to five resulted in a shift from transient coordinated to stably high expression states (Figures 5D and 5E, respectively). The shift from transient coordinated to stably high expression states is also reflected by the bimodal distribution of genes product counts for in the highly connected network (Figures 5F and 5G), where genes stay permanently in the high state once they leave the low expression state. These results mimic the experimentally measured mRNA expression states of the drug-induced reprogrammed melanoma cells.

To test if the computational prediction holds true in melanoma, we performed network inference using the φ-mixing coefficient-based (Ibragimov, 1962) Phixer algorithm (Singh et al., 2018) on the experimental data (Box 2; STAR Methods, section Comparative Network Inference; Table S2). Specifically, we used the Phixer algorithm on the mRNA counts obtained from FISH imaging data of marker genes in drug-naive cells and the resistant colonies that emerge post-drug treatment to infer the underlying network. Consistent with the model prediction, we found that the number of edge connections (for a range of edge weight thresholds) between marker genes increased substantially for 6 out of 7 resistant colonies compared with the drug-naive cells (Figure 5H). To control for biases from subsampling of the experimental data and nature of the Phixer algorithm itself (see STAR Methods, section Comparative Network Inference), we ran the entire network inference analysis 1,000 times. Again, in all 1,000 runs, we saw a higher number of total edges for 6 out of 7 resistant colonies than the drug-naive cells (Figures 5I, S8A, and S8C).

Besides the dependence on networks, our framework predicts that for a given network, stronger interactions between nodes (defined by the interaction parameter r_{add}) can also result in stable gene expression profiles (Figures S6E–S6G). It is possible that reprogramming results from a combination of increased edge connectivity as well as the enhanced interactions (given by parameter r_{add}) between existing edges. Biologically, it would translate into stronger and increased number of interactions between genes and associated transcription factors during reprogramming. Together, network inference of the experimental data is consistent with model findings about the cellular progression from a transient coordinated high expression state to a stably high expression state.

DISCUSSION

We developed a computational framework to model rare cell behaviors in the context of a drug-naive melanoma population where a rare subpopulation of cells displays transient and coordinated high gene expression states. We found that a relatively parsimonious stochastic model consisting of transcriptional bursting and stochastic interactions between genes in a network is capable of producing rare coordinated high states that mimic the experimental observations. To systematically investigate their origins, we screened networks of increasing sizes and

time-region (two-sample Kolmogorov-Smirnov test, p value < 0.001), implying that enhanced transcriptional activity is caused by more frequent bursts rather than prolonged bursts.

(D) Violin plots of the fold change in number of high states and total time spent in high states for network 3.2 and its unconnected graph. Positive regulatory interactions between the connected nodes (network) leads to an increased number of and total time in high states in comparison to independent nodes. Fold change is calculated by dividing the number of high states (total time spent in high states) for network 3.2 with the number of high states (total time spent in high states) for the unconnected graph. Each dot represents one of the 26 simulations showing rare coordinated high states for network 3.2.

(E) Distributions of burst duration in the baseline time-region (black) and those coincident with entry time-point (gray) (see Figure 4A). The bursts coincident with entry time-points are significantly longer than bursts in the baseline time-region (two-sample Kolmogorov-Smirnov test, p value < 0.001).

(F) Distributions of burst duration in the high time-region but not the exit time-region ([high-exit] time-region) (light gray) and those in the exit time-region (dark gray) (see Figure 4A). There is no statistically significant difference between the distributions underlying the duration of bursts in the high time-region and the exit time-region (two-sample Kolmogorov-Smirnov test, p value > 0.05).

(G) Violin plots of the mean burst duration ratios for entry and exit ($n_{entry} = n_{exit} = 594$), where mean burst ratio represents the difference in means of the burst duration distributions (see E and F) per simulation for all simulations with rare coordinated high states. Ratio close to 1 suggests no difference between the two regions. While the mean (and median) burst duration ratio between entry time-point and baseline time-region is considerably increased, the mean (and median) burst duration ratio between bursts in the exit time-region and in the rest of the high time-region are comparable for all simulations with rare coordinated high states.

(H and I) Distributions of the time intervals between genes entering (H) and exiting (I) the high time-region, denoted by t_{ent} and t_{exit} respectively in Figure 4A, are distributed differently for two representative simulations. While the time intervals for entering (t_{ent}) the high time-region are not exponentially distributed (H) (and hence not random), the time intervals for exiting (t_{exit}) the high time-region are exponentially distributed (I) (Lilliefors test, p value < 0.001 and > 0.05, respectively). See also Figure S7.

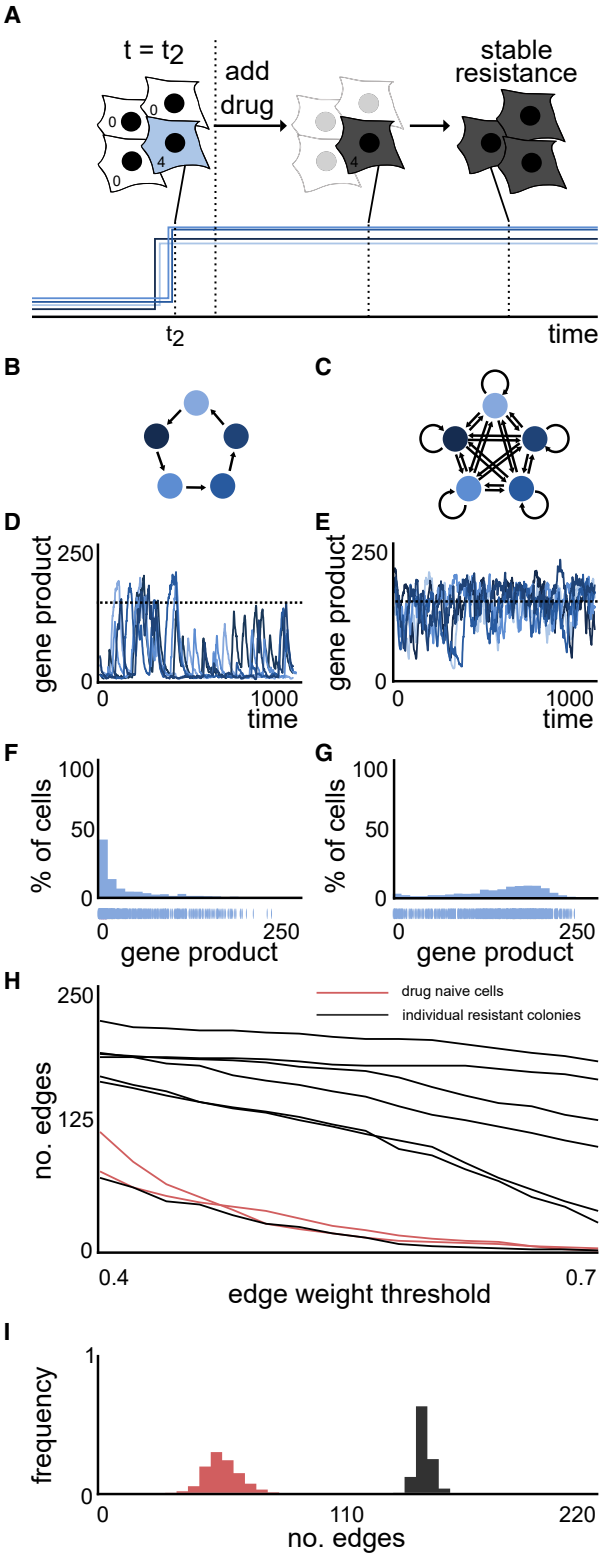


Figure 5. Increased Connectivity of a Network Leads to Stable High Expression Which Is Also Observed in Emerging Resistant Colonies Post-drug Treatment

(A) Upon drug treatment, the surviving cells acquire stable resistance. A schematic gene expression pattern is shown below.

(B–E) (B and C) Networks of size 5 with low (B) (1) and high (C) (5) connectivity and corresponding (D and E) simulations.

(F and G) The expression distributions are determined by taking the counts of simulated gene products per 1,000 time units (see Figure 2A) of simulations (D and E) corresponding to the lowly (B) and highly (C) connected networks. The gene expression distribution of the highly connected network (G) does not exhibit heavy tails while the simulation of the lowly connected network (F) exhibits heavy tails.

(H) Comparison of the connectedness of the underlying inferred gene regulatory networks of drug-naive cells and resistant colonies (post-drug treatment) using the Phixer algorithm for network inference analysis. Total number of edges is calculated for different edge weight thresholds, defined as the threshold at which an inferred edge is assumed to be present in the inferred gene regulatory network. For all the edge weights investigated, 6/7 resistant colonies have inferred gene regulatory networks with higher numbers of edges than drug-naive cells, suggesting that the gene regulatory networks underlying resistant colonies are more strongly connected.

(I) Applying the network inference analysis 1,000 times for a fixed edge weight threshold of 0.45 gives distributions for the number of edges in the inferred gene regulatory networks for both drug-naive cells (red) and resistant colonies (black) (distributions shown for one example each). The distribution of number of edges in the inferred gene regulatory network is considerably increased for the resistant colony.

See also Figure S8.

connectivities for a broad range of parameter values. Our study revealed that they occur more frequently for networks with low connectivity and depend on 3 of the 7 independent model parameters. Furthermore, we showed that the mechanisms that lead to the transition into and out of the rare coordinated high state are fundamentally different from each other. Collectively, our framework provides an excellent basis for further mechanistic and quantitative studies of the origins of rare, transient, and coordinated high expression states.

Given the relative generality of the networks that produce rare coordinated high states, the transcriptional bursting model predicts that every cell type is capable of entering the rare coordinated high state. Furthermore, we show that canonical modes of transcription alone, namely the binding of the transcription factor at gene locus to produce mRNA via recruitment of RNA polymerase II, can lead to these states without requiring other complex mechanisms, such as DNA methylation, histone modifications, or phase separation. Although such other mechanisms may still be operational in these cells to regulate their entry to or exit from these states, we posit that in principle, any set of genes interacting via traditional gene regulatory mechanisms are capable of exhibiting these rare coordinated high states, as long as they are interacting in a certain manner (e.g., sparsely connected) with appropriate kinetic parameters. In the case of drug-naive melanoma cells, the transient state is characterized by an increased ability to survive drug therapy leading to uncontrolled proliferation of the resulting resistant cells. It is possible that these rare transient behaviors may exist across many sets of interacting genes that may or may not manifest into phenotypic consequences. Another possibility the transcriptional bursting model predicts is that even within the same cell, distinct modules of interacting genes can lead to distinct sets of rare

coordinated high states that each can affect the cellular function and outcomes differently. These possibilities can be tested for by using increasingly accessible single cell RNA sequencing techniques on clonal populations of cells.

One limitation of the transcriptional bursting model is that we have performed quantitative analysis only on symmetric networks with positive interactions between nodes. Although the preliminary analysis on two cases of randomly selected asymmetric networks shows that they do exhibit the rare coordinated high states (Figures S2G–S4M), it remains to be seen whether these findings hold more generally for asymmetric networks. Inhibitory interactions between nodes are a separate and perhaps more interesting point. In principle, the model can be adapted to include inhibitory interactions. These inhibitory interactions may lead to non-monotonic effects of network connectivity on the occurrence of rare states, as positive and negative interactions can compete in non-linear ways. Similarly, a network with both negative and positive interactions may be more prone to instability, even for relatively smaller networks. Furthermore, inclusion of these interactions might also make the exit of genes from the high expression state dependent on one another, which occurs independently in the current transcriptional bursting model. We also highlight that unlike the experimental data, the model simulations do not have non-zero values for a larger number of genes in the high expression states (Figure 2B). The absence of non-zero values may be because the network underlying the experimental data contains a much larger set of interacting genes, thereby increasing the likelihood of non-zero values for a higher number of expressed genes. Larger gene networks can be explored in the future studies.

While we have focused on rare, transient, and coordinated high expression states in melanoma, our study provides conceptual insights into other biological contexts, such as stem cell reprogramming. Particularly, there is increasing evidence to suggest that stem cell reprogramming to desired cellular states proceeds via non-genetic mechanisms in a very rare subset of cells (Hanna et al., 2009; Pour et al., 2015; Takahashi and Yamanaka, 2016). The transcriptional bursting model may explain the origins and transient nature of this type of rare cell variability. In sum, we have established the plausibility that a relatively parsimonious model comprising of transcriptional bursting and stochastic interactions of genes organized within a network can give rise to a new class of biological heterogeneities. Therefore, we believe that established principles of transcription and gene expression dynamics may be sufficient to explain the extreme heterogeneities that are being reported increasingly in a variety of biological contexts.

Key Changes Prompted by Reviewer Comments

In response to the reviewers' comments, we made the introductory paragraph concise, added Box 1, which provides detailed description and associated assumptions of the model, and added Box 2, which provides definitions of metrics used to quantify the rare coordinated high states. We also relaxed the model assumptions (Figures S4A–S4E; STAR Methods) to explore the effect of (1) including translation and (2) using a multiplicative mode of gene interaction. Additionally, we performed extensive mechanistic analysis of the model features that initiate the transition into rare coordinated high states and those that enable maintenance of these states. Findings from this analysis are presented in Figure 4 and Results section. Furthermore, we analyzed additional network topologies (Figures 2 and S5), tested the model on a bigger network size (10 nodes) (Figure S2D), and performed sensitivity analysis on the parameter space (Figure S6D). We also performed comparative analysis between experimental data for multiple genes and computational data using two metrics (1) Gini coefficient measuring entropy (Figure 2D) and (2) fitting exponentials to analyze for sub-exponentiality (Figure S3C). For context, the complete transparent peer review record is included within the Supplemental Information.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- METHOD DETAILS
  - Networks
  - Models
  - Model Selection
  - Parameters
  - Simulations
  - Simulation Classes
  - Network Topologies
  - Quantitative Analysis
  - Decision Tree Optimization, Generalized Linear Models and Constrained Simulations
  - Sensitivity Analysis
  - Burst Analysis: Maintenance of Rare Coordinated High States
  - Entry and Exit Mechanisms
  - Comparative Network Inference
  - Asymmetric Networks or Parameter Sets
  - Asymmetric Parameter Sets
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY
  - Data
  - Code

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.03.004.

ACKNOWLEDGMENTS

We thank the Raj lab members, especially Ian Mellis and Amy Azaria, for scientific discussions and comments on the manuscript. We also thank Ravi Radhakrishnan and Alok Ghosh for helpful discussion during the initial stages of this project. We thank Cesar A Vargas-Garcia for his help during the initial discussions on network inference. L.S. would like to acknowledge the support of the PROMOS fellowship of the DAAD, Germany. L.S. was funded by the BMBF project TIDY (031L0170B) and financially supported by the Entrepreneurial Award within the program 'Global Challenges for Women in Math Science' of the Department of Mathematics, Technical University of Munich. B.L.E. acknowledges support from NIH F30 CA236129 and Patel Family Scholars award. A.S. acknowledges support from the NIH grant 5R01GM124446-02

and ARO grant W911NF-19-1-0243. C.M. acknowledges support from the Deutsche Forschungsgemeinschaft DFG through the SFB 1243. A.R. acknowledges support from NIH/NCI PSOC U54 CA193417, NSF CAREER 1350601, P30 CA016520, SPORE P50 CA174523, NIH U01 CA227550, NIH 4DN U01 HL129998, NIH Center for Photogenomics RM1 HG007743, NIH R01 CA232256, NIH R01 CA238237, NIH R01 GM137425, and the Tara Miller Foundation. Y.G. would like to acknowledge the Schmidt Science Fellows in partnership with the Rhodes Trust. Y.G. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research and this investigation has been aided by a grant from the Jane Coffin Childs Memorial Fund for Medical Research.

AUTHOR CONTRIBUTIONS

Conceptualization, A.R. and Y.G.; Methodology, L.S., Y.G., and A.R.; Software, L.S. and A.R.; Validation, L.S.; Formal Analysis, L.S. and M.S.-A.; Resources, A.R. and A.S.; Investigation, E.M.S., Y.G., and B.L.E.; Data Curation, L.S. and Y.G.; Writing – Original Draft, Y.G.; Writing – Review & Editing, A.R., L.S., Y.G., C.M., E.M.S., B.L.E., and M.S.-A.; Visualization, L.S. and Y.G.; Supervision, Y.G., A.R., and C.M.; Project Administration, Y.G. and A.R.; Funding Acquisition, A.R., A.S., and C.M.

DECLARATION OF INTERESTS

A.R. receives royalties related to Stellaris RNA FISH probes. All other authors declare no competing interests.

Received: July 16, 2019
Revised: February 3, 2020
Accepted: March 24, 2020
Published: April 22, 2020

REFERENCES

Antolović, V., Miermont, A., Corrigan, A.M., and Chubb, J.R. (2017). Generation of single-cell transcript variability by repression. Curr. Biol. *27*, 1811–1817.e3.

Breiman, L., Friedman, J., Stone, C.J., and Olshen, R.A. (1984). Classification and Regression Trees (Wadsworth Statistics/Probability), First Edition (Chapman & Hall/CRC).

Chen, H., and Larson, D.R. (2016). What have single-molecule studies taught us about gene expression? Genes Dev. *30*, 1796–1810.

Corrigan, A.M., Tunnacliffe, E., Cannon, D., and Chubb, J.R. (2016). A continuum model of transcriptional bursting. eLife *5*, e13051.

Fallahi-Sichani, M., Becker, V., Izar, B., Baker, G.J., Lin, J.R., Boswell, S.A., Shah, P., Rotem, A., Garraway, L.A., and Sorger, P.K. (2017). Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. Mol. Syst. Biol. *13*, 905.

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. *81*, 2340–2361.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025–1036.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Ham, L., Schnoerr, D., Brackston, R.D., and Stumpf, M.P.H. (2020). Exactly solvable models of stochastic gene expression. bioRxiv. https://doi.org/10.1101/2020.01.05.895359v1.

Ham, L., Brackston, R.D., and Stumpf, M.P.H. (2019). Extrinsic noise and heavy-tailed laws in gene expression. bioRxiv. https://doi.org/10.1101/623371v1.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Huynh-Thu, V.A., and Sanguinetti, G. (2019). Gene regulatory network inference: an introductory survey. In Gene Regulatory Networks: Methods and Protocols, G. Sanguinetti and V.A. Huynh-Thu, eds. (Springer), pp. 1–23.

Ibragimov, I.A. (1962). Some limit theorems for stationary processes. Theory Probab. Appl. *7*, 349–382.

Iyer-Biswas, S., Hayot, F., and Jayaprakash, C. (2009). Stochasticity of gene products from transcriptional pulsing. Phys Rev E Stat Nonlin Soft Matter Phys. *79*, 031911.

Jiang, L., Chen, H., Pinello, L., and Yuan, G.C. (2016). GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. Genome Biol. *17*, 144.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. *33*, 269–276.

Phillips, R., Belliveau, N.M., Chure, G., Garcia, H.G., Razo-Mejia, M., and Scholes, C. (2019). Figure 1 theory meets figure 2 experiments in the study of gene expression. Annu. Rev. Biophys. *48*, 121–163.

Pisco, A.O., and Huang, S. (2015). Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: "what does not kill me strengthens me". Br. J. Cancer *112*, 1725–1732.

Pour, M., Pilzer, I., Rosner, R., Smith, Z.D., Meissner, A., and Nachman, I. (2015). Epigenetic predisposition to reprogramming fates in somatic cells. EMBO Rep. *16*, 370–378.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Saint-Antoine, M.M., and Singh, A. (2019). Evaluating pruning methods in gene network inference. arXiv http://arxiv.org/abs/1902.06028.

Schlauch, D., Glass, K., Hersh, C.P., Silverman, E.K., and Quackenbush, J. (2017). Estimating drivers of cell state transitions using gene regulatory network models. BMC Syst. Biol. *11*, 139.

Shaffer, S.M., Emert, B.L., Reyes-Hueros, R., Coté, C., Harmange, G., Sizemore, A.E., Gupte, R., Torre, E., Singh, A., Bassett, D.S., and Raj, A. (2018). Memory sequencing reveals heritable single cell gene expression programs associated with distinct cellular behaviors. bioRxiv. https://doi.org/10.1101/379016.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435.

Sharma, A., Cao, E.Y., Kumar, V., Zhang, X., Leong, H.S., Wong, A.M.L., Ramakrishnan, N., Hakimullah, M., Teo, H.M.V., Chong, F.T., et al. (2018). Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. Nat. Commun. *9*, 4931.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69–80.

Singh, N., Ahsen, M.E., Challapalli, N., Kim, H., White, M.A., and Vidyasagar, M. (2018). Inferring genome-wide interaction networks using the phi-mixing coefficient, and applications to lung and breast cancer. IEEE Trans. Mol. Biol. Multi-Scale Commun. *4*, 123–139.

So, L.H., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. Nat. Genet. *43*, 554–560.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428–432.

Su, Y., Wei, W., Robert, L., Xue, M., Tsoi, J., Garcia-Diaz, A., Homet Moreno, B., Kim, J., Ng, R.H., Lee, J.W., et al. (2017). Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc. Natl. Acad. Sci. USA *114*, 13679–13684.

Symmons, O., and Raj, A. (2016). What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. Mol. Cell *62*, 788–802.

Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. *17*, 183–193.

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. USA *98*, 8614–8619.

Torre, E.A., Arai, E., Bayatpour, S., Beck, L.E., Emert, B.L., Shaffer, S.M., Mellis, I.A., Budinich, K.A., Weeraratna, A., Shi, J., and Raj, A. (2019). Genetic screening for single-cell variability modulators driving therapy resistance. bioRxiv. https://doi.org/10.1101/638809.

Van Kampen, N.G. (1992). Stochastic Processes in Physics and Chemistry (Elsevier).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
RNA-FISH data – marker genes	Shaffer et al., 2017	https://www.dropbox.com/sh/g9c84n2torx7nuk/AABZei_vVpcfTUNL7buAp8z-a?dl=0
RNA-FISH data – network inference (resistant colonies)	Shaffer et al., 2017	https://www.dropbox.com/sh/g9c84n2torx7nuk/AABZei_vVpcfTUNL7buAp8z-a?dl=0
Data – Model simulations	This paper	https://www.dropbox.com/sh/n94q45zkn5w54fe/AACC3cgts4kD6MWEE452pEgEa?dl=0
Software and Algorithms		
MATLAB R2017a and R2018a	Mathworks	https://www.mathworks.com
Phixer	Singh et al., 2018	https://github.com/nitinksingh/phixer/
Code – Model simulations	This paper	https://doi.org/10.5281/zenodo.3713697

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yogesh Goyal (yogesh.goyal0308@gmail.com). This study did not generate any new materials.

METHOD DETAILS

Networks

In our framework, the nodes in the network represent genes, where the expression of a gene is regulated by the expression of other genes. Gene regulation is represented by directed edges in the network, e.g. if the expression of gene Y is regulated by the expression of gene X, then the network contains an edge from node X to node Y. These networks can be defined by adjacency matrices given by:

$$A_{ij} = \begin{cases} 1, & \text{if } \textit{there is an edge from node } i \textit{ to } j \\ 0, & \textit{else}. \end{cases}$$

Any node in a network of size N can be connected with up to N-1 other nodes and in the case of self-loops, to N other nodes. Hence, the adjacency matrix A is of size N*N. This means that there are 2^{NxN} possible adjacency matrices for a network of size N - each of the possible N*N matrix entries can take on one of the values of 0 (no edge) and 1 (edge). For example a network of size 3 has $2^{(3*3)} = 512$ possible networks.

Here, we focus on symmetric networks, where we assume a relational identity between all nodes in a network. Experimental data from Shaffer et al. (Shaffer et al., 2017) implies the absence of any obvious hierarchical structure within the genes, and that the driver genes may interact in a relatively non-hierarchical manner (Figure S1D). The structural embedding of a node in its network can increase or decrease its ability of being involved in coordinated overexpression. For example, a centered node within a star-shaped network is involved more frequently in coordinated overexpression than the other nodes within the same network (Figure S1E), which is inconsistent with the experimental observations. To ensure for non-hierarchical behavior we define a set of symmetric networks (Figure S1F), where the number of in- and outgoing edges within a node and across nodes is identical and either all nodes in a network have a self-loop or not, leading to adjacency matrices of which the rows are cyclic permutations (to the right) with offset one of each other. We first compute all possible vectors $\{0,1\}^N$, in total 2^N vectors. From each of these resulting vectors, we create an NxN matrix by using the given (row) vector as template, and creating the other N-1 rows by cycling the prior row vector to the right by one step, where the right-most entry in the row vector is added to the (so far empty) left-most entry. By applying this permutation N-1 times, all possible cyclic permutations are captured within a matrix, and each node in the given network is completely relational identical. We make use of the *circshift* function in MATLAB to receive the possible cyclic permutations of the initial row vectors.

We further constrain the analysis to weakly-connected networks – any node in a network has to be connected to at least one other node, without taking into account the directionality of the edges. In terms of the adjacency matrix:

$$\forall i \in \{1, \ldots, N\} : \sum_{j \in \{1,\ldots,N\},\, j \neq i} A_{ij} + A_{ji} \geq 1.$$

The above restriction allows us to exclude the consideration of compositions of smaller and unconnected networks, which could otherwise lead to double counting. These subnetworks of smaller sizes are analyzed in the sets of networks of respective node sizes. To perform this operation, we analyze all the previously constructed adjacency matrices using the MATLAB function *conncomp(X,'Type','weak')*, which assigns each node with a bin number according to the connected component of its underlying undirected graph. If all nodes of a network belong to the same bin number i.e. to the same connected component, the adjacency matrix encodes for a weakly-connected graph. Finally, we further restrict the analysis to non-isomorphic networks. Two networks are called isomorphic if there exists a bijection from the edge space of one network to the other, such that any edge of one network is projected to a particular edge in the other network. Here, the labeling of the nodes (gene 1, gene 2,...) in the networks is arbitrary and hence relabeling of nodes in an adequate fashion leads to identical networks. To ensure that all the final networks analyzed are of a non-isomorphic set of networks, we test all networks with MATLAB's function *isisomorphic*. We initiate the final set of networks with one adjacency matrix, and then sequentially test all other networks for isomorphism. If the given network is non-isomorphic to the current final set, it is added to the final set. Conversely, if the network is isomorphic to one of the networks in the final set, it is discarded.

By reducing the possible set to weakly-connected, non-isomorphic and symmetric networks, we greatly reduce the possible number of networks. For example, in the previous example, we had 512 possible networks for 3 nodes. By applying all the mentioned constraints (weakly-connected, non-isomorphic and symmetric), 4 networks remain (Figure S1C). We perform the analysis on networks of sizes 2, 3, 5 and 8 each consisting of 2, 4,10 and 80 networks, respectively, adding up to a total of 96 networks (Figure S9). In principle, the transcriptional bursting model can easily be extended to larger network sizes without the loss of generality (Figures S2D–S2F).

Models

Model 2 - Transcriptional Bursting Model

The transcriptional bursting model is an expansion of the telegraph model, where DNA can take on one of the two states, active and inactive, e.g. based on the presence or absence of transcription factors (Figure 1C). The active and inactive state directly translates into high and low rates of production of gene products, respectively. We add interaction terms to the model, where the expression of a gene influences the rate of DNA activation of another gene depending on how they are organized in a respective network. Here we use the number of mRNA as a faithful proxy for the number of proteins. In other words, we only model the number of mRNA counts and assume that any mRNA is immediately translated into one single functional protein after its translation. Therefore, the mRNA count determines the strength of the regulation. Here, we model the regulation of one gene by another using the Hill function, given by:

$$f(\mathrm{mRNA}_X) = \frac{\mathrm{mRNA}_X^n}{k^n + \mathrm{mRNA}_X^n},$$

where mRNA_X is the mRNA count of gene X, n is the Hill coefficient and k is the dissociation constant, n,k > 0. The Hill coefficient determines the steepness of the Hill function, i.e., the extremeness of its switch-like effect. The dissociation constant determines the half-maximal value, $f(\mathrm{mRNA}_X) = 0.5$.

The reversible transitions between the inactive and active states, as well as the mRNA synthesis and degradation, are modeled by chemical reactions. For each gene, we have three chemical species - the DNA inactive state, the DNA active state and mRNA. These three species interact with one another according to the following 5 chemical reactions:

$$I \rightarrow A$$

$$A \rightarrow I$$

$$I \rightarrow I + \mathrm{mRNA}$$

$$A \rightarrow A + \mathrm{mRNA}$$

$$\mathrm{mRNA} \rightarrow \varnothing,$$

defining the corresponding stoichiometric matrix:

$$\begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 \end{pmatrix}.$$

The stoichiometric matrix encodes the net change in each chemical species resulting from any of the chemical reactions where the chemical reactions are assumed to occur stochastically. Under the assumptions of the law of mass action, the probability of a specific molecular collision to occur in the infinitesimal time interval [t, t + dt) is proportional to the product of the molecule counts of the

educt chemical species. The reaction propensity $a_j(x)$ for a given chemical reaction R_j and state x, determines the probability density function such that $a_j(x)dt$ gives the probability of the chemical reaction R_j taking place in dt, for small dt. Examples of reaction propensities for so called elementary reactions are given here:

Reaction	Reaction propensity
$\varnothing \rightarrow$ products	k
$X_i \rightarrow$ products	kx_i
$X_i + X_j \rightarrow$ products	kx_ix_j

where k is called the reaction rate.
The gene regulation influences the reaction rate of the DNA activating chemical reaction.
To explain the above-mentioned chemical reactions, we introduce eight rates/parameters:

Parameter	Description
r_{on}	The rate at which DNA is activated.
r_{off}	The rate at which DNA is inactivated.
r_{prod}	Synthesis rate of mRNA.
r_{deg}	Degradation rate of mRNA.
r_{add}	Parameter determining the contribution of the additional DNA activation rate upon gene regulation.
d	Factor by which the mRNA synthesis rate is increased when in an active DNA state (in comparison to basal synthesis rate in DNA inactive state), >1.
k	Dissociation constant of the Hill function.
n	Hill coefficient.

The full model description for one gene regulated by a single gene X is given below:

Chemical reaction	Reaction rate	Reaction propensity
$I \rightarrow A$	$r_{on} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}$	$\left(r_{on} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}\right) \cdot I$
$A \rightarrow I$	r_{off}	$r_{off} \cdot A$
$I \rightarrow I + mRNA$	r_{prod}	$r_{prod} \cdot I$
$A \rightarrow A + mRNA$	$d \cdot r_{prod}$	$d \cdot r_{prod} \cdot A$
$mRNA \rightarrow \varnothing$	r_{deg}	$r_{deg} \cdot mRNA$

where $I, A \in \{0, 1\}$, and $I + A = 1$, where $I = 0$ ($A = 1$) denotes that the DNA is in an active state and $I = 1$ ($A = 0$) denotes that the DNA is in an inactive state. $mRNA_X$ is the mRNA count of gene X at the given time, r_{on} is the basal DNA activation rate, r_{add} is the additional activation rate due to gene regulation, r_{off} is the DNA inactivation rate, r_{prod} is the basal mRNA synthesis rate in the DNA inactive state, d denotes the increase in the mRNA synthesis rate when the DNA is in the active state, where $d > 1$, and r_{deg} is the mRNA degradation rate. The chemical reactions are identical for all N nodes in a given network of size N. The reaction rate of activation ($I \rightarrow A$), composed of terms with parameters r_{on} and r_{add}, is the only node-specific rate. It depends on the underlying network and has to be adapted accordingly for each node, where the in-going edges of a node determine which gene regulations are active. The addition of hill function-based activation terms corresponds to the adaptation of the standard telegraph model. We model gene regulation additively: if there is more than one influencing gene, we add the Hill function terms of the respective genes. As an example, if the gene of interest is influenced not only by gene X, but by gene X and gene Y, the activation rate from above will expand to:

$$r_{on} + r_{add} \cdot \left(\frac{mRNA_X{}^n}{k^n + mRNA_X{}^n} + \frac{mRNA_Y{}^n}{k^n + mRNA_Y{}^n} \right).$$

We also tested for multiplicative regulation, i.e. regulation where we multiply the reaction rates (and consequently the reaction propensities) of the influencing genes (Figure S4C). In the example above the activation rate then expands to

$$r_{on} + r_{add} \cdot 2 \cdot \left(\frac{mRNA_X^n}{k^n + mRNA_X^n} \cdot \frac{mRNA_Y^n}{k^n + mRNA_Y^n} \right)$$

instead. By definition the Hill function is restricted to values between 0 and 1. While a multiplication of two Hill functions results in a maximal value of 1, an addition results in a maximal value of 2. As the Hill function is an important factor in these simulations we hence add a scaling factor to the activation rate in case of multiplicative regulation. We show that for network 5.3, 97 out of 1000 simulations show rare coordinated high states in case of multiplicative regulation (Figures S2D and S2E). In comparison, 15 simulations show rare coordinated high states in case of additive regulation. 9 simulations show rare coordinated high states in both cases.

Additionally, we tested for translation events (Figure S4A). We added one state (P) and two rate parameters, a protein synthesis rate r_{prodP} and a protein degradation rate r_{degP}, to the original transcriptional bursting model. The extended model description accounting for translation for one gene regulated by gene X is given below:

Chemical reaction	Reaction rate	Reaction propensity
$I \rightarrow A$	$r_{on} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}$	$\left(r_{on} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}\right) \cdot I$
$A \rightarrow I$	r_{off}	$r_{off} \cdot A$
$I \rightarrow I + mRNA$	r_{prod}	$r_{prod} \cdot I$
$A \rightarrow A + mRNA$	$d \cdot r_{prod}$	$d \cdot r_{prod} \cdot A$
$mRNA \rightarrow \varnothing$	r_{deg}	$r_{deg} \cdot mRNA$
$mRNA \rightarrow mRNA + P$	r_{prodP}	$r_{prodP} \cdot mRNA$
$P \rightarrow \varnothing$	r_{degP}	$r_{deg} \cdot P$

where we define k again as 0.95 of the high steady state, this time for the protein count:

$$k(r_{prodP}, r_{degP}, d, r_{prod}, r_{deg}) = 0.95 \cdot \frac{r_{prodP}}{r_{degP}} \cdot d \cdot \frac{r_{prod}}{r_{deg}},$$

which itself is dependent on the high steady state of the mRNA (d * rprod/rdeg). Redefining $r_{prodP} = a * r_{prod}$ and $r_{degP} = b * r_{deg}$ gives

$$k(r_{prod}, r_{deg}, d) = 0.95 \cdot d \cdot \frac{a}{b} \cdot r_{prod}^2 \cdot r_{deg}^2.$$

We tested three different translation scenarios: protein synthesis and degradation being (1) faster than (2) same as and (3) slower than mRNA synthesis and degradation. For network 5.3 and parameter set 968, giving rise to rare coordinated high states in the transcriptional bursting model without translation, we took a = b = 10 (faster), a = b = 1 (same) and a = b = 0.1 (slower) as additional parameters. We find that protein synthesis and degradation with faster (Figure S4B) and same rates as mRNA degradation and synthesis, also allows for the formation of rare coordinated high states in the case of translation. Only slower protein synthesis and degradation rates did not show rare coordinated high states, likely because for faster protein rates, the system dynamics is determined largely by the transcriptional dynamics. In sum, we demonstrate that the rare coordinated high states can arise in the revised model that includes translation.

Model 1- Constitutive Model

Model 1 is a simple gene regulatory expression model, where mRNA can either be transcribed or degraded and the mRNA of a regulatory gene influences the transcription rate of a regulated gene (Figure 1B). Here again, we assume the number of mRNA to be a faithful proxy for the protein number and hence, only model the mRNA expression of a gene. The gene regulation is modeled according to the Hill function (STAR Methods; Model 2 - Transcriptional Bursting Model).

The synthesis and degradation are modeled by chemical reactions. For each gene, we have one chemical species, its mRNA, described by the following two chemical reactions:

$$\varnothing \rightarrow mRNA$$

$$mRNA \rightarrow \varnothing,$$

defining the corresponding stoichiometric matrix:

$$(1 \quad -1).$$

The full model description for one gene regulated by a single gene X is given below:

Chemical reaction	Reaction rate	Reaction propensity
$\varnothing \rightarrow mRNA$	$r_{prod} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}$	$r_{prod} + r_{add} \cdot \frac{mRNA_X^n}{k^n + mRNA_X^n}$
$mRNA \rightarrow \varnothing$	r_{deg}	$r_{deg} \cdot mRNA$

where r_{prod} the basal mRNA synthesis rate, r_{deg} the mRNA degradation rate, r_{add} the additional synthesis rate due to gene regulation and $mRNA_X$ the mRNA count of gene X at the given time.

The chemical reactions are identical for all N nodes in a given network of size N. The synthesis rate is a node-specific rate (STAR Methods; Model 2 - Transcriptional Bursting Model). We model gene regulation additively (STAR Methods; Model 2 - Transcriptional Bursting Model). For k we tested two different definitions: one closer and one further away from the low expression taking into account the intrinsic stochasticity. We therefore first run a test simulation with a random k for 1,000 time units and determine the standard deviation of the expression of the node denoted as 'node 1'. k is latin hypercube sampled with the rest of the parameters with lower and upper boundary 100 and 1000. We set k to be:

$$k = \frac{r_{prod}}{r_{deg}} + x \cdot std,$$

where std is the standard deviation of the expression of the node denoted as 'node 1' and $x \in \{3,5\}$. We then re-initiate the simulation with the adapted k value.

Model Selection

We decided to develop a network-based framework that models the cell-intrinsic biochemical interactions. One of the first goals we had was to identify the minimal set of biochemical reactions that constitutes this network model. We asked whether a simple network model lacking gene activation step (Model 1), i.e. with constitutive mode of gene expression, is sufficient to capture rare coordinated high states (Figure 1B; STAR Methods; section Model 1)? Or that we need to incorporate gene activation step via transcriptional bursting (Model 2) at each node, a phenomenon in which genes flip reversibly between transcriptionally active and inactive state regulated by the binding of a transcription factor(s) (Figure 1C; STAR Methods; section Model 2)?

In terms of chemical reactions, the critical difference between the two models is that, while in Model 1 the gene is transcribed as a Poisson process with a single rate, r_{prod} (Figure 1B), in Model 2, a gene can reversibly switch between active (r_{on}) and inactive state (r_{off}), where binding of the transcription factor at a gene locus defines the effective rate of gene production (Figure 1C). Specifically, when inactive, the gene is transcribed as a Poisson process at a basal rate (r_{prod}); when active, this rate becomes higher (d x r_{prod}, where $d > 1$). For both the models, we modeled degradation of the gene product as a Poisson process with degradation rate r_{deg}. For both the models, the inter-node interaction parameter, r_{add}, has a Hill-function-based dependency on the gene product amount (Hill coefficient n) of the respective regulating node to account for the multistep nature of the interaction (Figures 1B and 1C). All chemical reactions, propensities, and model parameters are presented in STAR Methods. To test these two models, we used Gillespie's next reaction method (Gillespie, 1977) and simulated test cases of small networks (of two or three nodes) for a range of parameters.

For a vast majority of the networks and parameter combinations, Model 1 either produced always low or always high expression states (Figure S1A). In some cases, while Model 1 could indeed produce a transition from low to high expression states, the transition happens for all gene products at the same time (Figure S1A). However, this model is not consistent with the experimental observations; in particular, if a cell is positive for one marker gene, then it is more likely to be positive for another marker gene, but not necessarily so (Figure S1B) (Shaffer et al., 2017). Furthermore, this mode of transition resulted in bimodal distributions of cellular state as determined by the amount of gene product (Figure S1B), which is different from the rare nature of the transitions, as reflected by the heavy-tailed distributions of gene products observed in melanoma. Model 2, which incorporates transcriptional bursting-dependent activation of a node (gene), also produced a range of gene expression states (Figures 1C–1F). Importantly, this model was able to faithfully capture the qualitative features of the experimental data i.e. rare, transient, and coordinated high expression states (Figure 1F). In contrast to Model 1, Model 2 captures another property of the experimental data, i.e. if one gene is in the high expression state, the other genes in the network are likely to be in high expression state, but not always (Figures 2B and S2B). Based on these initial observations, we decided to pursue Model 2 systematically and simulated networks of different sizes and architectures across a broad range of model parameters.

Parameters

The goal of our study is to model the emergence of rare transient coordinated high expression of several genes. The theoretical idea behind the transcriptional bursting model is that each time the DNA is in an active state, corresponding to a transcriptional burst, the steady-state of the mRNA count is shifted from r_{prod}/r_{deg} to d^*r_{prod}/r_{deg}. Accordingly, the mRNA attempts to reach its new steady-state which results in a rapid increase in their counts. Depending on the length of the transcriptional burst, which is exponentially distributed with rate parameter r_{off}, the mRNA count is able to reach the new steady-state. We use the dynamical system behavior

when modeling the rare coordinated overexpression. In principle, for most transcriptional bursts, the sudden mRNA increase should not initiate a DNA activation of its regulated genes; only in some rare cases, the transcriptional burst in one gene is long enough such that its mRNA count exceeds a certain threshold that may be able to affect the state of another gene locus on DNA. Exceeding of the mRNA threshold can lead to an increased probability of the DNA states of its regulated genes to be activated and hence to an increased mRNA synthesis in the respective genes. The increased mRNA synthesis of regulated genes may lead to positive feedback loops network-wide resulting in the transient coordinated overexpression of genes.

The threshold to be overcome by the mRNA count of a gene to make its gene regulation effective is given by the dissociation constant of the Hill function, k. k determines the 'switching point' from (almost) no gene regulation to (almost) complete gene regulation. Therefore, we define k to be a function of r_{prod}, r_{deg} and d as follows:

$$k\left(r_{prod}, r_{deg}, d\right) = 0.95 \cdot d \cdot \frac{r_{prod}}{r_{deg}},$$

where d^*r_{prod}/r_{deg} gives the steady-state mRNA count of the respective regulating gene in the DNA active state. Here, we arbitrarily determine the threshold k to 0.95 of its high-expression steady-state to restrict the emergence of coordinated overexpression to being rare and for the system to demonstrate a considerable difference between the low and high gene expression state. The simulations and the analysis are all performed according to the above definition of k. We tested the robustness of this definition for a particular network 5.3 (Figure S9) where we performed the same simulations (for 100 latin hypercube sampled parameter sets (Table S1)) as for the final analysis as before using five different definitions of k:

$$k\left(r_{prod}, r_{deg}, d\right) = x \cdot d \cdot \frac{r_{prod}}{r_{deg}},$$

where $x \in \{0.75, 0.8, 0.85, 0.9, 1\}$ (Table S1). Our analysis shows that for x = 0.75, none of the 100 simulations show rare coordinated high states: the threshold leading to an effective gene regulation is exceeded too often: the regulated DNA states are activated, the high state emerges and we lose the rareness of the coordinated high gene expression event. The number of simulations showing rare coordinated high states increases with increasing x, reaching its maximum for x = 0.95 (standard, 7 out of the 100 simulations show rare behavior). For x = 1 (high expression steady-state), we also see rare behavior in 7 out of 100 simulations, showing overlapping results in 6 out of the 7 simulations.

Together, we are left with a set of seven parameters consisting of: r_{on}, r_{add}, n, r_{off}, r_{prod}, d, r_{deg}, which may be split into inter-gene (r_{on}, r_{off}, r_{prod}, d, r_{deg}) and intra-gene (r_{add}, n) parameters and the dependent parameter k. Potentially, these parameter sets are node-dependent resulting in a N * 7-dimensional parameter space for a network of size N.

To emphasize the equality between the nodes, we use the same 7-dimensional parameter set for all nodes in a network. Hence, the nodes are relationally and parametrically identical, thereby also allowing us to directly compare the simulations of different network sizes, otherwise not possible, and to determine the effects of network size and architecture on the ability of forming the rare coordinated high state. Therefore, we latin-hypercube sample 1000 parameter sets out of the parameter space with upper and lower boundaries (chosen arbitrarily, but typically spanning two orders of magnitude):

Parameter	Lower boundary	Upper boundary
r_{prod}	0.01	1
r_{deg}	0.001	0.1
r_{on}	0.001	0.1
r_{off}	0.01	0.1
d	2	100
r_{add}	0.1	1
n	0.1	10

by using the MATLAB function *lhsdesign_modified* (Khaled, N. Latin Hypercube (https://de.mathworks.com/matlabcentral/fileexchange/45793-latin-hypercube), MATLAB Central File Exchange. Retrieved May 5, 2018.). The 1000 parameter sets are shown in the Table S1. For some plots, we used a y-axis break function in MATLAB (Mike, C.F. Break Y Axis (https://www.mathworks.com/matlabcentral/fileexchange/45760-break-y-axis), MATLAB Central File Exchange. Retrieved December 21, 2018.)

Simulations

We simulated model 2 for a total of 96 networks (for all weakly-connected, non-isomorphic, symmetric networks of sizes 2, 3, 5 and 8 with 2, 4, 10 and 80 networks, respectively)(Figure S9), each for 1,000 sampled parameter sets, resulting in a total of 96,000 simulations across four different network sizes. The simulations were performed according to Gillespie's next reaction method and were computed for 1,000,000 time units, which is critical for capturing rare behaviors. For all simulations, the DNA state was initiated (t = 0) to be in its inactive state and the mRNA count was arbitrarily set to 20 for all nodes. The mRNA counts quickly reach their low-expression steady state, such that we are certain that our analysis is not impaired by the given initial conditions. The simulations were

implemented in MATLAB R2017a and R2018a. One single simulation of 1,000,000 time units took between 20 minutes and 9 hours depending on the parameter set and the network. The complete simulations took over 1.5 months to run, where we parallelised all 96 networks and let each of them run on four cores simultaneously.

Simulation Classes

We analyzed all of the 96,000 simulations, and assign them to the following four classes, initially by visual inspection, and subsequently by defined criteria (see below):

I - stably low gene expression
II - stably high gene expression
III - uncoordinated transient high gene expression
IV- rare, transient coordinated high gene expression

Therefore we constructed three criteria, for which all the simulations were tested. We primarily focus on the rare, transient coordinated high gene expression states, as defined by the following criteria:

1) Coordinated high gene expression state. We call a simulation to show coordinated high expression, if at least once within the 1,000,000 time unit simulation more than half of the mRNA counts are above a specified threshold (e.g. for 5 nodes, at least once three or more mRNA counts have to be above a defined threshold; for 8 nodes, at least once 5 or more mRNA counts have to be above a defined threshold). Similar to the definition of the dissociation constant k, we set the threshold to

$$\text{thres} = 0.8 \cdot d \cdot \frac{r_{prod}}{r_{deg}},$$

where $d * r_{prod}/r_{deg}$ gives the high-expression steady state. Again, we want to detect the rare occurrence of a large mRNA count deviation from the low-steady state and hence, set the threshold arbitrarily to 0.8 (see below for details on the choice of this value).

To compare the simulated results with the experimental data from a drug-naive melanoma cell population, we split the 1,000,000 time unit simulations into 1,000 time unit sub-simulations, each accounting for a cell. Hence, we receive simulations of 1,000 cells for 1,000 time units, a procedure justified by the ergodic theory. To show that sub-simulations of 1,000 time units are uncorrelated, we determine the autocorrelations for all 1,000 parameter sets of network 3.2 (Figure S9) for up to 1,000 lags (using the MATLAB autocorrelation function *acf* (Autocorrelation function(ACF) (https://www.mathworks.com/matlabcentral/fileexchange/30540-autocorrelation-function-acf), MATLAB Central File Exchange. Retrieved June 13, 2019.). For each of these, we determine the first lag at which the autocorrelation is below the upper 95% confidence bound. For 88.2% of all simulations, the first lag below the upper 95% confidence bound occurs before 1,000 lags. For the 26 simulations with rare coordinated high states, 23 show a first lag below the upper 95% confidence bound before 1,000 lags. For the remaining three simulations the autocorrelation after 1,000 lags is at 0.0615, 0.0206 and 0.4363. Removing the simulation with high autocorrelation (0.4363) does not change the conclusions of our analysis.

2) Rareness/transience. To mimic the results given by RNA-FISH in a drug naive melanoma population, where we only see a snapshot of the mRNA counts within a melanoma cell, we randomly determine a time-point t_{rand}, where $t_{rand} \in [0,999]$ (uniformly distributed), at which we count the number of mRNA counts above the threshold (for each simulation t varies). We summarize the result of all 1,000 cells in a histogram, for which we expect a decrease with increasing mRNA count above the threshold.
3) Heavy-tailed gene expression distributions. At the population level, the single mRNA distributions of marker genes show heavy-tails. We use the same time-point t as sampled for criterion 2) and consider the mRNA counts of all genes. If we plot these in gene-dependent histograms, we expect to find right-skewed and unimodal distributions. Here, we use the MATLAB function *skewness(X)* for evaluating the right-skewness of the histogram, where skewness(X) > 0, denotes that the data is spread out more to the right of the mean. Skewness is defined as

$$\text{skewness}(X) = E\left[\frac{(X-\mu)^3}{\sigma^3}\right]$$

where μ is the mean of X, σ is the standard deviation of X and E(.) the expectation. For determining unimodality, we test whether the maximum of the last quarter of histogram bins with bin width of one is less than the minimum of the first quarter of histogram bins. Although the definition above only characterizes a heavy-tailed distribution, we find it to be sufficient for our analysis.

Classes I and III, are both defined by criterion 1 only, where criterion 1 is not met in both cases. For class I, none of the genes in a network ever express above the given threshold. For class III, genes express above the given threshold but not once are more than half of the genes above the given threshold at any given time of the simulation. Only if a simulation is able to fulfill all three criteria, will we call it a simulation of class IV - rare transient coordinated high gene expression. If a simulation fulfills criteria 1, but fails to meet both other criteria, we classify it into class II. To receive numbers of simulations in class IV - rare transient coordinated high expression - per network size, we randomly determine three different t_{rand}, where each $t_{rand} \in [0,999]$ (uniformly distributed) and evaluate all

96000 simulations for being in class IV at the respective snapshot (Figure 2A). Note that all these requirements are tested automatically using a script without manual/human intervention.

To show that criterion 3) is sufficient for defining heavy-tailed simulations in class IV in our analysis, we constrain criterion 3) further aiming to identify sub-exponentially decaying, heavy-tailed distributions more directly. We therefore reevaluate all simulations so far identified as class IV and compare their 99th percentiles of their expression distributions with those of fitted exponential distributions (Figure S3C, right panel). We expect most of the 99th percentile of the expression distributions to be larger than the 99th percentile of the fitted exponentials. Due to the symmetry of the networks and the resulting similarity between the expression distributions (Figure S2C), we only consider node one here, without the loss of generality. To avoid that the fitted exponentials account for the heavy-tails, we constrain the fits to have a maximal bin number (bin size of one) within $\mp$ 1 of the maximal bin number (bin size one) of the expression distributions. We do so by sequentially increasing/decreasing the exponential parameter μ by steps of 10, sampling 1000 times from the resulting exponential distribution with the MATLAB function *exprnd*(μ,1,1000) and comparing the maximal bin number of the resulting histograms. We repeat the above until the maximal bin number of the exponential distribution is within the predefined range of $\mp$ 1. As expression distributions with a large maximum bin are more similar to lognormal distributions with small variances and less to exponentials, we restrict the analysis to expression distributions with a maximum bin of $\leq$ 15 (Figure S3B). The threshold of a maximum bin of 15 was determined by considering the simulations and their exponential fits. We additionally discard simulations for which the optimization takes more than 1000 iterations or is producing non-positive parameter values.

Most (82%) of the 99th percentile of the simulated expression distributions are above the diagonal, hence larger than the 99th percentile of the fitted exponential distributions (Figure S3C, right panel). The 99th percentile of all the nine marker genes in Shaffer et al. also lie above the diagonal in the general vicinity of the points corresponding to simulations with rare coordinated high states (Figure S3C, left panel). We therefore conclude that criterion 3) sufficiently selects for sub-exponentially decaying heavy-tailed distributions.

We additionally, perform parts of the analysis again on two different levels of stricter stringency for criterion of heavy-tailed distributions (Figures S4F–S4M):

A) All simulations fulfilling criteria 1) - 3) which additionally comply to the above mentioned analysis (maximum bin $\leq$ 15, 99th percentile of expression distribution > 99th percentile of fitted exponential, <1000 iterations to reach a $\mp$ 1 of the maximal bin number (bin size one) in the optimization for determining the exponential fit and producing non-positive parameter values) (Figures S4J–S4M)
B) All simulations fulfilling criteria 1) - 3) which additionally comply to the above mentioned analysis or have a maximum bin > 15 (Figures S4F–S4I)

The results are qualitatively very similar to the results we receive if we perform the analysis only on criteria 1) - 3) (Figures 2, 3 and S4). The 6 and 7 rare coordinated high parameter sets identified by the more stringent analyses A) and B), respectively, are subsets of the original eight rare coordinated high parameter sets (Figures 3A, S4H, and S4L). Although the resulting optimized decision trees vary slightly, they still identify all three parameters, r_{on}, r_{add} and r_{off}, controlling rare transient coordinated states, as in the original analysis. Together, we conclude that the simple characterization of heavy-tailed distributions is sufficient for further analysis.

The analysis above is a prerequisite for further findings and statements. Due to its importance, we tested its robustness with respect to the definition of the threshold, marking the mRNA count above which a gene is called to be in the high-gene expression state, and with respect to the number of mRNA counts required above the threshold to call it a coordinated high state (both determining criterion 1).

For the test network 5.3, we hence repeated the analysis for thresholds:

$$\text{thres} = x \cdot d \cdot \frac{r_{prod}}{r_{deg}},$$

where x = 0.3 : 0.05 : 1 (here, for 100 latin hypercube sampled parameter sets (Table S1), and we only test for class IV). Decreasing the threshold down to 0.6 of the high-expression steady state does not change the set of simulations with rare behavior in comparison to the results for x = 0.8. Even a further decrease of the threshold (down to 0.3 of the high-expression steady state) manifests in a similar result: half of the simulations identified previously to show rare behavior are still classified as such. Hence, we keep x = 0.8 for the rest of the analysis (Table S1).

Next, for network 5.3 and the 100 parameter sets (Table S1), we repeated the analysis requiring at least 1, 2, 4, and 5 mRNA counts to be above the threshold at least once, in order for the simulation to fulfill criterion 1. The lower the required mRNA count, the more simulations fulfill criterion 1 (peaking at a required mRNA count of at least 1 with 11 out of the 100 simulations showing rare behavior according to this definition). The above set of simulations entails the set of simulations fulfilling criterion 1 at the standard required mRNA count of at least 3 (7 out of 100 simulations). Hence, we keep the definition of coordinated overexpression to more than half the nodes being above the threshold.

Additionally, we computed the Gini indices for the gene expression distributions of both the simulations showing rare coordinated high states and the experimental data (Figures 2D and S3A) (Jiang et al., 2016; Shaffer et al., 2017). A Gini coefficient of 0 implies perfect equality such that for a given gene, all cells within a population have the same number of mRNA molecules, whereas 1 implies perfect inequality such that one cell expresses all the mRNA molecules while others express none. We used the MATLAB function *gini* (Gini coefficient and the Lorentz curve (https://www.mathworks.com/matlabcentral/fileexchange/28080-gini-coefficient-and-the-lorentz-curve), MATLAB Central File Exchange. Retrieved October 24, 2019.) for the computations.

Network Topologies

Connectivity

We define a measure for the connectivity of the networks, where

connectivity = number of ingoing edges for any node of the network

where a self-loop is also considered to be an ingoing edge. As we constrain our analysis to symmetric networks (same number of ingoing edges for all nodes in a network per definition), we are able to define one single connectivity per network. The constraints enable us to directly evaluate the impact of the connectivity of the network on the ability to form rare behavior.

Self-loops

A network with a direct auto-activation is called a network with a self-loop. Due to the restriction of symmetric networks, all networks can be classified as having self-loops for all nodes or not having self-loop for any node. Due to non-isomorphism, the set of networks contains for each network without self-loops an identical network with self-loops. We evaluate the ability of these different edge classes on the formation of rare coordinated high states (Figure 2G).

Characteristic Distance

The characteristic distance of a network is defined as the average shortest path length for all pairs of nodes within a given network. To calculate this distance, we used the MATLAB function *shortestpath* on all pairs of nodes. We evaluated the ability of the characteristic distance normalized to the network size on the formation of rare coordinated high states (Figure S5F).

Quantitative Analysis

For each of the 96,000 simulations showing rare coordinated high states we performed a quantitative analysis. First, we define a high expression region as a region which is initiated by the first mRNA count to exceed the threshold, terminated by the last mRNA count to drop below the threshold and requires to contain a coordinated high expression state (criterion 1: more than half the mRNA counts have to exceed the defined threshold) between the initiation and termination time-points. Breaks of up to 50 time unit intervals are accepted due to the stochastic nature of the simulations. For example, in a 3 node network, where we require at least 2 mRNA counts to exceed the threshold for a coordinated high state: the first mRNA count exceeds the threshold (initiation), then the second mRNA count exceeds the threshold (initiation of high state) but then drops below the threshold for 50 time units before exceeding the threshold again, is still counted as one high-expression region. The length of 50 time units were defined arbitrarily. Due to the stochasticity of the system and the conservative definition of the threshold (located close to the high-expression steady state), we observe these temporary violations of criterion 1. In order to create sensible statistics on the quantitative behavior of the simulations, the temporary relaxation of criterion 1 is necessary.

In the quantitative analysis we extract the total time spent in a high state (out of 1,000,000 time units) from all simulations showing rare behavior (Figure S3D).

Decision Tree Optimization, Generalized Linear Models and Constrained Simulations

We classify all parameter sets into two classes, rare coordinated high parameter sets and non-rare coordinated high parameter sets, according to the percentage of total simulations per parameter set (96 simulations) in which rare coordinated high states are observed. The threshold above which a parameter set is called a rare coordinated high parameter set is at 20%. More than 19 of the 96 simulations have to show rare behavior in order for a parameter set to be called a rare coordinated high parameter set. The threshold was set according to a summarizing histogram, in which we see a clear distinction between the two groups: the main body of the histogram being located below 20% and the few parameter sets deviating extremely from that main group (> 20%). According to this binary classification, we performed a decision tree optimization (MATLAB function fitctree).

To validate the results of the decision tree optimization, we used generalized linear models on all seven independent parameters r_{on}, r_{add}, n, r_{off}, r_{prod}, d and r_{deg} with the MATLAB function fitglm(X,Y,'Distribution','binomial').

To validate that the parameter region determined by the decision tree optimization favors the formation of simulations with rare coordinated high states, we generate a new set of parameters constrained to values close to the minimal and maximal values of r_{on}, r_{add} and r_{off} for the rare coordinated high parameter sets:

Parameter	Lower boundary	Upper boundary
r_{prod}	0.01	1
r_{deg}	0.001	0.1
r_{on}	0.001	0.025
r_{off}	0.06	0.1
d	2	100
r_{add}	0.15	0.36
n	0.1	10

We latin hypercube sample 1000 parameter sets from that constrained parameter space. For all 1000 parameter sets we simulate 1000000 time units by Gillespie's next reaction method for networks 3.2 and 5.3 (Figure S9). Each of these simulations was evaluated for having rare coordinated high states according to the three criteria (STAR Methods, section Simulation Classes).

Sensitivity Analysis

For each parameter, we tested its sensitivity across its corresponding parameter space (see STAR Methods, section Parameters). Briefly, we take network 3.2 (Figure S9) for the detailed analysis as network 3.2 shows rare coordinated high states in all eight rare coordinated high parameter sets. For each of the seven independent parameters (r_{on}, r_{off}, r_{pod}, r_{deg}, n, d, r_{add}), we determine 10 equidistant points across its parameter space, and create new parameter sets by swapping these new parameters one-by-one with ones from the eight rare coordinated high parameter sets, resulting in 8*7*10 = 560 new parameter sets. We simulate 1,000,000 time units with Gillespie's next reaction method for these newly created parameter sets and evaluate all new simulations for showing rare coordinated high states. For each of the 10 newly sampled parameter values per parameter we receive 8 binary decisions where '1' indicates that the simulation exhibits rare coordinated high states and '0' that it does not. Our analysis confirmed that the three parameters (r_{on}, r_{off}, and r_{add}) identified by the decision tree algorithm and generalized linear model are indeed critical for producing the rare coordinated high states (Figure S6D). We also found a moderate dependence on the Hill coefficient n, also confirmed by the low p-value for n from generalized linear model analysis (Figure S6C).

Burst Analysis: Maintenance of Rare Coordinated High States

For all simulations showing rare coordinated high states, we determine the fraction and frequency of transcriptional bursts in both the high and baseline time-regions (Figures 4B and 4C). By fraction we mean the percentage of the total time the system is bursting. By frequency we mean the number of bursts per unit time. Additionally, we determine the number of high states and the total time spent in a high state for a network of size three (network 3.2, Figure S9) and three independent nodes for each of the parameter sets showing rare coordinated high states in the connected network (Figure 4D).

Entry and Exit Mechanisms

Entering/Exiting of High Expression Region - Transcriptional Bursts

For all of the simulations in class IV showing rare coordinated high states - we analyze whether the durations of transcriptional bursts are coordinated with the entering and exiting of high time-regions (Figure 4A; STAR Methods; section Quantitative Analysis).

For all of the simulations showing rare transient coordinated high gene expression, we analyze the distributions of waiting times between genes entering and exiting the high expression region (see Quantitative Analysis).

Entering high expression regions - For all high expression regions, we determine the first time-points at which the gene counts exceed the threshold (only for genes with a gene count exceeding the threshold during a particular high expression region at least once). We then consider the waiting times - the time interval between the ascending sorted time-points of genes entering the high expression region. These distributions - at most N-1 distributions for a network of size N, one for each waiting time between the genes - are compared to exponential distributions by the Lilliefors test according to the MATLAB function *lillietest(X, 'Distr', 'exp')* at a significance level of 0.05.

Entering/Exiting of High Expression Region – Times

For all of the simulations showing rare transient coordinated high gene expression, we analyze the distributions of waiting times between genes entering and exiting the high expression region (see Quantitative Analysis).

Entering high expression regions - For all high expression regions, we determine the first time-points at which the gene counts exceed the threshold (only for genes with a gene count exceeding the threshold during a particular high expression region at least once). We then consider the waiting times - the time interval between the ascending sorted time-points of genes entering the high expression region. These distributions - at most N-1 distributions for a network of size N, one for each waiting time between the genes - are compared to exponential distributions by the Lilliefors test according to the MATLAB function *lillietest(X, 'Distr', 'exp')* at a significance level of 0.05.

Exiting high expression regions - For all high expression regions we determine the last time-points at which the gene counts exceed the threshold (again, only for genes with a gene count exceeding the threshold during a particular high expression region at least once). We consider the waiting times and compare their distributions to exponential distributions by the Lilliefors test by applying the MATLAB function *lillietest(X, 'Distr', 'exp')* at a significance level of 0.05.

Comparative Network Inference

Here we describe the computational techniques we used to infer the gene interaction network structure of the pre-drug and post-drug cells. When studying regulatory interactions between genes in a network, it can be useful to abstract the problem into a graph theory framework. Let us assume a set of N genes, with the expression level of each gene represented by the random variable X_i, with $i \in \{1,...,N\}$. The network of interactions between genes can then be represented as a graph of N nodes. An edge $X_i \rightarrow X_j$ signifies a regulatory relationship in which X_i either upregulates or downregulates X_j (Singh et al., 2018).

The computational challenge of network inference is to uncover the true edges of the gene interaction network from statistical relationships between gene expression levels. Many different algorithms, often based on mutual information, conditional probability, or regression analysis, have been developed (Singh et al., 2018; Huynh-Thu and Sanguinetti, 2019; Saint-Antoine and Singh, 2019). The

output of an inference algorithm is a matrix of edge weights, which we will call W with dimensions NxN. In this matrix, the element w_{ij} is a measure of how confident we can be that the edge $X_i \rightarrow X_j$ exists in the network. A final network prediction will typically set a threshold for edge weights, and exclude any edges that fall below the threshold. Edges $X_i \rightarrow X_i$, called "self-edges" are typically excluded for the final network prediction, except in cases when temporal data is being analyzed. Since we are using atemporal expression data here, self-edges will be excluded from the analysis.

It is common to judge a network inference algorithm's reliability by testing it on a "gold standard" dataset, for which the true structure of the network is already known, to see how well it can recover the real edges from the expression data (Huynh-Thu and Sanguinetti, 2019). We have chosen to use the Phixer algorithm (Singh et al., 2018), based on its impressive performance when benchmarked on the DREAM5 Challenge gold standard datasets (weblink: http://dreamchallenges.org/project/dream-5-network-inference-challenge/; last accessed: 05/06/2019).

Phixer

Phixer computes edge weights using the phi-mixing coefficient. For discrete random variables X and Y taking values in sets A and B, the phi-mixing coefficient $\varphi(X|Y)$ is defined as:

$$\phi(X_i|X_j) = \max_{S\subseteq A, T\subseteq B}\left|\Pr\{X_i \in S | X_j \in T\} - \Pr\{X_i \in S\}\right|. \quad \text{(Equation 1)}$$

We then assign $\varphi(X_i|X_j)$ as the weight of the edge $X_j \rightarrow X_i$. The phi-mixing coefficient is an asymmetric measure, so the weight of the edge $X_i \rightarrow X_j$ may be different (Singh et al., 2018).

The original Phixer algorithm includes a pruning step, which attempts to correct for false positives by minimizing redundancy in the network. For every possible triplet of nodes X_i, X_j, and X_k, the following inequality is checked:

$$\varphi(X_i|X_k) \leq \min\{\varphi(X_i|X_j), \varphi(X_j|X_k)\} \quad \text{(Equation 2)}$$

If Equation 2 holds, the edge $X_k \rightarrow X_i$ is eliminated. However, previous work has found that the pruning step, though theoretically sensible, typically reduces accuracy in practice (Saint-Antoine and Singh, 2019), possibly due to the prevalence of redundant connections, such as feed forward loops in gene regulatory networks. So, we removed this part of the algorithm in order to achieve the highest possible level of accuracy.

The Phixer software is available online at the creator's GitHub page: https://github.com/nitinksingh/phixer/ (last accessed: 05/06/2019). We used the original C code, and kept the default parameter values the same, except for changing "NROW" to 19 and "TSAMPLE" to 4000, to reflect the dimensions of the input data files. The original Phixer code includes, by default, 10 bootstrapping runs, as well as a built-in procedure for binning the raw data, which we did not alter. We removed the pruning step from the code, but otherwise left the edge weight calculation process unchanged.

Data Description

The two pre-drug datasets are referred to as NoDrug1 and NoDrug2 in the supplementary data files (Table S2). The datasets containing clusters of resistant cells after four weeks of drug exposure are referred to as Fourweeks1-cluster1, Fourweeks1-cluster2, etc. where we differentiate between Fourweeks1 with four clusters and Fourweeks2 with three clusters. Details of how these datasets were acquired are presented in (Shaffer et al., 2017).

Bootstrapping Controls

We found that the Phixer algorithm tends to predict more connections for larger sample sizes, even when the samples are taken from the same dataset. To control for the differences in original sample sizes of various samples, we bootstrapped the original datasets into 4000-sample datasets before performing the Phixer analysis. The number 4000 was chosen arbitrarily; bootstrapped sample sizes of 1000, 2000, and 6000 also appeared to produce similar results.

Randomized Controls

For each size-controlled dataset to be analyzed, we created a randomized control consisting of permutations of each gene column from the original dataset (Table S2). We then performed the Phixer analysis on these randomized controls. The resulting edge weight distributions give us a baseline or control edge weight for Phixer that, in principle, reflects potential false positives. We found that in the controls, nearly all of the predicted edge weights were below 0.45 (Figure S8B). Therefore, we decided to choose 0.45 as a threshold for the non-control analysis, thus eliminating edges that could have been predicted by chance alone.

Finally, since the analysis contains two stochastic elements (the bootstrapping to correct for the sample size issue and the bootstrapping step in the Phixer algorithm itself) we had to be sure that the observed differences in connectivity were not due to chance. For each dataset, we ran the entire analysis (including both the bootstrapping size correction and the Phixer algorithm) 1000 times, and provide the distributions of the number of edges with weight greater than 0.45 (Table S2).

Asymmetric Networks or Parameter Sets

To test the generality of the results, we generate asymmetric simulations. We introduce asymmetry in both network architectures and the parameter sets.

Asymmetric Network

We randomly determine a weakly-connected but asymmetric five-node network (Figure S2G). We simulate the network with 100 parameter sets which are latin hypercube sampled out of the same parameter space as the 1000 parameter sets of the main analysis.

Out of these 100 simulations, two simulations are classified as showing rare, transient coordinated high gene expression (fulfills all three criteria in STAR Methods, section Simulation Classes, Figures S2H and S2I).

Asymmetric Parameter Sets

For the main analysis, we use the same parameter set, consisting of seven independent parameters (STAR Methods, section Parameters), for all nodes in a network. We introduce asymmetry by assigning each node in a network a separate set of parameters. Hence, we latin-hypercube sample 100 parameter sets out of a 7 x N parameter space, where N is the number of nodes of the network, with the MATLAB function *lhsdesign*_modified. Due to the high dimensionality, we here confine the parameter space to:

Parameter	Lower boundary	Upper boundary
r_{prod}	0.01	1
r_{deg}	0.001	0.1
r_{on}	0.001	0.1
r_{off}	0.001	0.1
d	2	100
r_{add}	0.2	0.4
n	5	10

where the changes in the boundaries are highlighted in blue. We confine the parameter space according to the clustering of rare coordinated high parameter sets. In total, six parameter sets give rise to rare-states more frequently than others for all 96 networks. Only two out of the seven independent parameters, r_{add} and n, show a strong correlation with the rare coordinated high state producing parameter sets as determined by a decision tree optimization. The boundaries in the table above are formed according to these decision tree boundaries in which five out of the six rare coordinated high state producing parameters lie (Table S1).

For these 100 parameter sets, we generated simulations for five-node network 5.3 (Figure S2J). Out of the resulting 100 simulations, we find two showing rare, transient coordinated high gene expression (fulfills all three criteria in STAR Methods, section Simulation Classes; Figures S2K–S2M).

QUANTIFICATION AND STATISTICAL ANALYSIS

Figure 2E: Independent sampling of t_{rand} was performed 3 times. Boxplots show the median and 25th and 75th percentiles. Figures 4B, 4C, 4E, and 4F: Two-sample Kolmogorov-Smirnov test tested for significance level 0.05. Figures 4H and 4I: Lilliefors test tested for significance level 0.05. Figures S4F and S4J: Independent sampling of t_{rand} was performed 3 times. Boxplots show the median and 25th and 75th percentiles. Figures S5A and S5B: Independent sampling of t_{rand} was performed 3 times. Boxplots show the median and 25th and 75th percentiles. Figures S7A and S7B: Lilliefors test tested for significance level 0.05.

DATA AND CODE AVAILABILITY

Data

The data used and generated in this manuscript is available via Dropbox (https://www.dropbox.com/sh/n94q45zkn5w54fe/AACC3cgts4kD6MWEE452pEgEa?dl=0).

Code

The MATLAB code used for the analysis of this manuscript is available on GitHub and the DOI is accessible via Zenodo (https://doi.org/10.5281/zenodo.3713697). The analysis was performed with MATLAB R2017a and R2018a.

A.2. H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested Xenopus embryos.

H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested Xenopus embryos.

Lea Schuh, Carolin Loos, Daniil Pokrovsky, Axel Imhof, Ralph A.W. Rupp, and Carsten Marr.

Histone modifications such as methylation alter the chromatin accessibility of genes thereby regulating their expression. Particularly during development, where cells continuously divide and differentiate, a fast and economical control of gene expression is required. However, cells divide rapidly during early development. With each cell cycle newly formed, largely unmodified histones are incorporated into the DNA leading to an overall dilution of most histone modifications. How is the histone modification landscape shaped by the cell cycle in vivo?

To identify whether the cell cycle plays an active role in shaping the histone 4 lysine 20 methylation (H4K20me) landscape, we compared the H4K20me kinetics between cycling and cell-cycle-arrested Xenopus embryos. First, we formulated a set of plausible mathematical models by ordinary differential equations to explain the H4K20me kinetics in cycling Xenopus embryos. We accounted for dilution of H4K20me proportions upon DNA replication by integrating and testing three different cell cycle functions. Allowing for shared and specific methylation and demethylation rate constants in addition to different cell cycle functions we received a set of 180 model hypotheses to explain the H4K20me kinetics in cycling Xenopus embryos. Next, we performed multi-start maximum likelihood optimization to infer the model parameters of all models and calculated the Bayesian Information Criterion for each of the optimized models for model selection which reduced the set of plausible models from 180 to 12. We found that only one of the three cell cycle functions (constrained scaled Hill function with Hill coefficient 1 and offset 0.5) was able to reproduce a biologically meaningful average cell cycle duration of 8 hours. While methylation rate constants were necessary to be specific for each methylation step, demethylation rate constants were found to be shared. Additionally, we found that models without demethylation performed just as well as models with demethylation suggesting that demethylation is not required to explain the H4K20me kinetics in cycling Xenopus embryos. We validated our findings by comparing the inferred cell cycle duration of our top performing model with cell cycle durations of Xenopus neural progenitors reported in literature. Furthermore, we used our top performing model to predict the effect of morpholino knockdowns of the di- and tri-methyltransferases which matched the experimental data assuming a 90% knockdown efficacy. We repeated the multi-start maximum likelihood optimization and model selection on all 30 possible models describing the H4K20me kinetics in cell-cycle-arrested Xenopus embryos. Similar to our findings in cycling embryos, we found methylation to be specific while demethylation to be shared. However, for cell-cycle-arrested Xenopus embryos the presence of demethylation was necessary to explain the data. Finally, we devised a joint model taking into account H4K20me kinetics of cycling and cell-cycle-arrested Xenopus embryos simultaneously. Constraining the joint model to our previous findings, we evaluated 40 possible models on the joint H4K20me kinetics. Performing model selection, we found a subset of 6 plausible models which all required a similar set of cycling-specific rate constants. By considering the marginal distributions of the inferred rate constants, we found that our model identified known cell-cycle-dependent rates. Furthermore,

our results suggest that demethylation is only required in cell-cycle-arrested Xenopus embryos likely serving as an alternative mode for regulation of H4K20me in cell-cycle-arrested Xenopus embryos to dilution in cycling Xenopus embryos.

Statement of individual contribution

This work was motivated by a discussion between Ralph Rupp, Axel Imhof, Carsten Marr and me during which we questioned whether the cell cycle plays a mere passive role in shaping the histone modification landscape during embryogenesis. I subsequently formulated a mathematical model describing the histone methylation kinetics for both cycling and cell-cycle-arrested Xenopus embryos, identified a subset of plausible models by optimizing the mathematical formulation based on ordinary differential equations on a time-series mass spectrometry data set of H4K20 methylation [49] and identified joint and cycling-specific rate constants. I was responsible for the entire methodology, formal analysis and the writing of the manuscript. Merely, the subsection *Experimental model and subject details* in *STAR METHODS* describing the experimental work was written by Daniil Pokrovsky.
I, Lea Schuh, am the main author of this publication.

Permission to include:

Schuh, L., Loos, C., Pokrovsky, D., Imhof, A., Rupp, R.A.W., Marr, C. (2020). H4K20 methylation is differently regulated by dilution and demethylation in proliferating and cell-cycle-arrested Xenopus embryos. *Cell Systems*, *11*(6), 653-662. doi: https://doi.org/10.1016/j.cels.2020.11.003

The following page contains a copy of the first page of the editorial policies of Cell Systems, including the author's rights. Here is stated, that 'as an author you may [...] include the article in full length or in part in a thesis or dissertation (provided that this is not to be published commercially)' (accessed February 2021). A digital version can be found here: https://www.cell.com/trends/editorial-policies

February 2021, https://www.cell.com/trends/editorial-policies

Editorial Policies

(…)

Authors' Rights

As an author, you (or your employer or institution) may do the following:

- Make copies (print or electronic) of the article for your own personal use, including for your own classroom teaching use.
- Make copies and distribute such copies (including through e-mail) of the article to known research colleagues for the personal use by such colleagues (but not for commercial purposes, as described below).
- Post a revised personal version of the final text (including illustrations and tables) of the article (to reflect changes made in the peer review and editing process) on your personal or institutional website or server, with a link (through the relevant DOI) to the article as published, provided that such postings are not for commercial purposes, as described below. Please note: Depositing in or posting to special repositories (such as PubMed Central or institutional repositories) is permitted only under specific agreements between Elsevier and the repository and only when consistent with Elsevier's policies concerning such repositories.
- Present the article at a meeting or conference and distribute copies of the article to the delegates attending such meeting.
- Allow your employer to use all or part of the information in the article for other intracompany use (e.g., training) if the article is a "work for hire" made within the scope of your employment.
- Retain patent and trademark rights and rights to any process or procedure described in the article.
- Include the article in full or in part in a thesis or dissertation (provided that this is not to be published commercially).
- Use the article or any part thereof in a printed compilation of your works, such as collected writings or lecture notes (subsequent to publication of the article in the journal).
- Prepare other derivative works that extend the article into book-length form or otherwise re-use portions or excerpts in other works, with full acknowledgment of its original publication in the journal.

All copies (print or electronic) or other use of the paper or article must include the appropriate bibliographic citation for the article's publication in the journal. However, you should not indicate in the citation that the version that you are reproducing or posting is the final published version as published in the journal. For example, it may be appropriate to indicate, "This paper has been submitted to [Journal] for consideration."

Commercial purposes include: the posting by companies or their employees for use by customers (e.g., pharmaceutical companies and physician prescribers); commercial exploitation such as associating advertising with such posting (including the linking to advertising by search engines); the charging of fees for document delivery or access; or the systematic distribution to others via e-mail lists or list servers (to parties other than known colleagues), whether for a fee or for free.

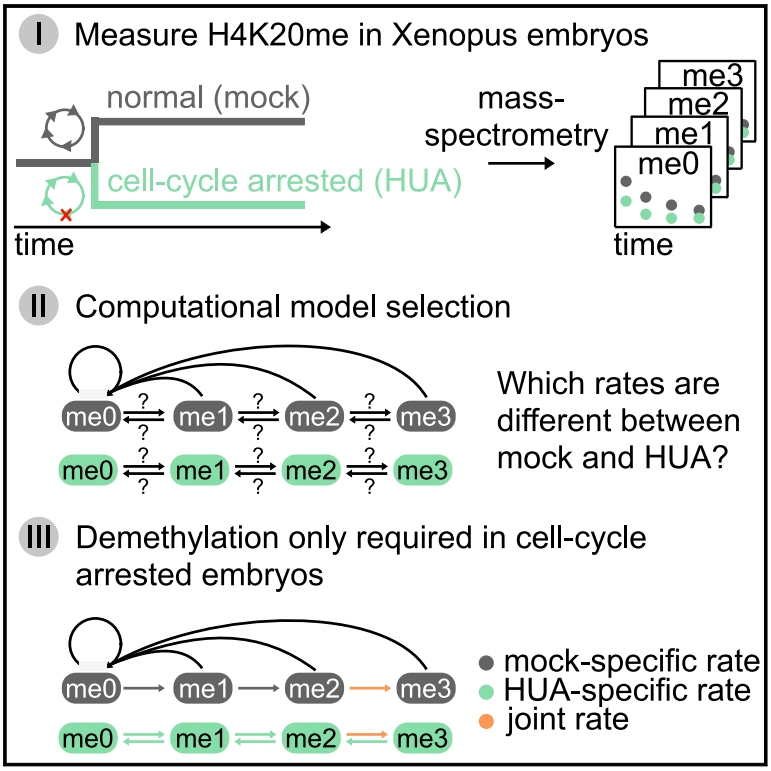
Cell Systems Editorial Policies Version: February 3, 2021

Cell Systems

Report

H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos

Graphical Abstract



Highlights

- Computational model explains H4K20me kinetics in cycling and arrested embryos
- Specific methylation rates are necessary—unspecific demethylation seems sufficient
- Active demethylation is required in cell-cycle-arrested embryos only
- Cell cycle may actively shape the H4K20me landscape during embryogenesis *in vivo*

Authors

Lea Schuh, Carolin Loos, Daniil Pokrovsky, Axel Imhof, Ralph A.W. Rupp, Carsten Marr

Correspondence

carsten.marr@helmholtz-muenchen.de

In Brief

Schuh et al. introduce a computational model to describe H4K20me kinetics in normal and cell-cycle-arrested Xenopus embryos. This quantitative model invokes specific methylation and unspecific demethylation and correctly predicts cell-cycle durations and cell-cycle dependencies. Active demethylation is not required to explain H4K20me kinetics of cycling cells, suggesting that overall H4K20me dilution through DNA replication is dominant. So only once cells stop cycling during embryogenesis, active H4K20 demethylation may contribute to shape histone methylation.

Schuh et al., 2020, Cell Systems *11*, 1–10
December 16, 2020
https://doi.org/10.1016/j.cels.2020.11.003

CellPress

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

Cell Systems

CellPress
OPEN ACCESS

Report

H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos

Lea Schuh,[1,2] Carolin Loos,[1,2,3,4] Daniil Pokrovsky,[5] Axel Imhof,[5] Ralph A.W. Rupp,[5] and Carsten Marr[1,6,*]
[1]Institute of Computational Biology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg 85764, Germany
[2]Department of Mathematics, Technical University of Munich, Garching 85748, Germany
[3]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA
[4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Department of Molecular Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, 82152, Germany
[6]Lead Contact
*Correspondence: carsten.marr@helmholtz-muenchen.de
https://doi.org/10.1016/j.cels.2020.11.003

SUMMARY

DNA replication during cell division leads to dilution of histone modifications and can thus affect chromatin-mediated gene regulation, raising the question of how the cell-cycle shapes the histone modification landscape, particularly during embryogenesis. We tackled this problem by manipulating the cell cycle during early *Xenopus laevis* embryogenesis and analyzing *in vivo* histone H4K20 methylation kinetics. The global distribution of un-, mono-, di-, and tri-methylated histone H4K20 was measured by mass spectrometry in normal and cell-cycle-arrested embryos over time. Using multi-start maximum likelihood optimization and quantitative model selection, we found that three specific biological methylation rate constants were required to explain the measured H4K20 methylation state kinetics. While demethylation is essential for regulating H4K20 methylation kinetics in non-cycling cells, demethylation is very likely dispensable in rapidly dividing cells of early embryos, suggesting that cell-cycle-mediated dilution of H4K20 methylation is an essential regulatory component for shaping its epigenetic landscape during early development.
A record of this paper's transparent peer review process is included in the Supplemental Information.

INTRODUCTION

All cells in our body contain the same genetic information encoded in the DNA. However, we are constituted out of many different cell types all performing their own specialized functions. Chromatin, mainly composed of DNA and histone octamers (two copies of histone H2A, H2B, H3, and H4 each), is an instructive DNA scaffold that aids extracting cell-specific information for gene expression. Histone tails are subject to various post-translational modifications, such as methylation, acetylation, phosphorylation, and ubiquitination (Bannister and Kouzarides, 2011), which play a fundamental role in altering chromatin accessibility. Dynamic regulation of gene expression is central for executing cell internal programs (proliferation, differentiation, etc.) and reacting to cell external signals with an appropriate response. Particularly during development, where cells continuously divide and differentiate, a fast and economical control of gene expression is required. Histone modifications are believed to regulate the progression throughout development (Jambhekar et al., 2020). In *Xenopus laevis*, a model organism for developmental biology, stage-specific histone modifications have been observed during the transit from pluripotent to differentiated states, a process called epigenome maturation (Schneider et al., 2011). However, cells divide rapidly during early development. With each cell cycle newly formed, largely unmodified histones are incorporated into the DNA leading to an overall dilution of most histone modifications (Jasencakova et al., 2010). How is the histone modification landscape shaped by the cell cycle *in vivo*?

Histone methylation is known to play important roles in many biological processes (Greer and Shi, 2012), and its deregulation is linked to cancer and aging in humans (Fraga et al., 2005; Klutstein et al., 2016). The methylation of lysine 20 on histone H4 (H4K20) is one of the most frequent lysine methylation sites observed in HeLa cells, mouse embryonic fibroblasts and several other cell types (Evertts et al., 2013; Leroy et al., 2013; Pesavento et al., 2008; Schotta et al., 2008). It is evolutionarily conserved from *Schizosaccharomyces pombe* to humans (Lachner et al., 2004) and is known to have a strong cell-cycle dependence. H4K20 occurs in four different states, un-, mono-, di-, and tri-methylation. Each methylation state plays a different functional role ranging from DNA-damage repair and chromatin condensation observed in fission yeast *Schizosaccharomyces pombe* (Sanders et al.,

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

2004), over transcriptional regulation shown in human T cells (Barski et al., 2007) and Xenopus embryos (Nicetto et al., 2013), mitotic progression found in *Drosophila melanogaster* (Sakaguchi and Steward, 2007), to cell-cycle control (Schotta et al., 2008), and silencing of repetitive DNA and transposons observed in mouse models (Schotta et al., 2004) and Xenopus embryos (van Kruijsbergen et al., 2017). H4K20me is regulated by three methyltransferases: KMT5A (also known as PR-Set7) for mono-methylation, first identified in *Drosophila* (Fang et al., 2002; Nishioka et al., 2002; Xiao et al., 2005), and SUV4-20H1 and SUV4-20H2 for both di- and tri-methylation, first identified in mammalian cells (Schotta et al., 2004). Whether there is a specificity of SUV4-20H1/2 for di- or tri-methylation is still debated (Schotta et al., 2008). The level of mono-methyltransferase KMT5A is cell-cycle dependent, and its degradation in G1 phase leads to a decline of H4K20me1 in late G1 as observed in human cell lines and Xenopus egg extracts (Abbas et al., 2010; Centore et al., 2010; Zee et al., 2012). H4K20me1 reaches its lowest level in S phase while increasing in G2 phase and peaking during mitosis. Both H4K20me2 and H4K20me3 levels have also been found to be cell-cycle dependent in HeLa cells though in a less dramatic fashion (Pesavento et al., 2008). The cell-cycle-dependent presence of H4K20 methyltransferases allows H4K20me2 and H4K20me3 to be reestablished only after mitosis in the next cell cycle (Jørgensen et al., 2013). For demethylation, unspecific enzymes such as PHF8 have been observed in human cell lines (Feng et al., 2010), but their functional importance has recently been questioned (Alabert et al., 2020; Jørgensen et al., 2013; Reverón-Gómez et al., 2018). It has even been suggested that the loss of histone mark H3K27me3 in mammalian cells may occur only by dilution during chromatin replication rather than by active removal (Jadhav et al., 2020). Finally, homologs of all H4K20-modifying enzymes are present in the Xenopus genome (Bowes et al., 2010).

To address the role of the cell cycle for epigenome maturation in Xenopus development, we have measured histone modification proportions in sibling embryo populations, which either proliferate or are arrested at the G1/S transition. Using quantitative mass spectrometry data for H4K20 we compared over 200 model hypotheses describing H4K20me kinetics in the cycling and cell-cycle-arrested population. With only a few assumptions, our computational model is able to explain H4K20me kinetics, retrieves correct cell-cycle durations and known cell-cycle dependencies of H4K20me. Furthermore, our approach allows us to estimate cell numbers over time and reveals the importance of three specific biological methylation rate constants and a shared biological demethylation rate constant, which is essential to establish the observed histone modification profile in the cell-cycle arrested but not required in the cycling population of Xenopus embryos.

RESULTS

Cell-Cycle Arrest Changes H4K20me Patterns during Xenopus Embryogenesis

After *in vitro* fertilization of a Xenopus oocyte, cells rapidly divide in a state of transcriptional quiescence up to 5.5 h post fertilization (hpf) (Heasman, 2006). Only then a regular zygotic cell cycle containing G1 and G2 phases is initiated (Newport and Kirschner, 1982). To identify how H4K20 methylation (H4K20me) is shaped by cell-cycle, we compared a population of normal Xenopus embryos (from now on called "mock") with a cell-cycle-arrested population. For this, half of the embryos were continuously incubated with hydroxyurea/aphidicolin (from now on called "HUA") from gastrulation onward (11 hpf). This treatment arrests cells at the G1/S boundary and is compatible with embryonic development (Harris and Hartenstein, 1991). HUA treatment applied before 11 hpf is lethal (Harris and Hartenstein, 1991; Pokrovsky et al., 2020). Correct and robust establishment of the cell-cycle arrest by HUA in the Xenopus embryos has been shown by Pokrovsky et al. (2020). Mass spectrometry measurements of H4K20me states, averaging over all cells in the embryos and all histones in the cells, were conducted at 14.75, 19.75, 27.5, and 40 hpf corresponding to late gastrula (NF13), neurula (NF18), tailbud (NF25), and tadpole (NF32) stages, respectively (Figure 1A). H4K20me proportions of mock and HUA showed significant differences across three biological replicates in all four H4K20me states (Figure 1B). In HUA-treated embryos, methylation accumulates in the di- and tri-methylation states in comparison to mock. Upon DNA replication, newly synthesized and unmodified histones are incorporated in mock, while in HUA, only little DNA replication takes place and hence only little incorporation of newly synthesized and unmodified histones occurs. All three biological replicates result in highly reproducible H4K20me proportions across all four developmental stages suggesting a high accuracy and quality of the mass spectrometry measurements.

Specific Methylation Rate Constants Are Necessary to Explain H4K20me in Mock Embryogenesis While Demethylation Is Not Essential

To identify how H4K20me kinetics are shaped by cell cycle, we defined models for untreated embryos (mock) and fitted them to the data (see STAR Methods). Mock models are composed of four H4K20me states corresponding to un- (me0), mono- (me1), di- (me2), and tri-methylated (me3) H4K20, allowing for successive methylation and demethylation with biological rate constants m_i and d_i, $i \in \{1,2,3\}$, respectively (see Figure 2A and STAR Methods for a detailed model description). For mock, where the cells undergo cell division, newly synthesized and unmethylated histones are incorporated into replicating DNA leading to a continuous dilution of methylated H4K20. Considering methylation proportions (defined as the frequency of a particular methylation state divided by the sum of all methylation states as measured by mass spectrometry), cell-cycle results in an overall increase of unmethylated H4K20 mediated by an outflow of H4K20me states with population growth rate $g(t) = \ln(2)/c(t)$, where $c(t)$ is the average cell-cycle duration c across all cells as a function of experiment time t (see Figure 2A and STAR Methods). Having measured average H4K20me proportions across whole Xenopus embryos, our cell-cycle function accordingly models average cell-cycle durations across all cells constituting the Xenopus embryos at the respective developmental stages. By considering average H4K20me proportions across asynchronous cell populations (Boterenbrood et al., 1983), we assume H4K20me dilution to occur continuously. The most general model is parameterized with six biological rate constants, where a biological rate constant is defined as the proportion of H4K20 in a particular methylation state being methylated/demethylated per hour (h^{-1}). Although no actual enzymatic rate

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

CellPress
OPEN ACCESS

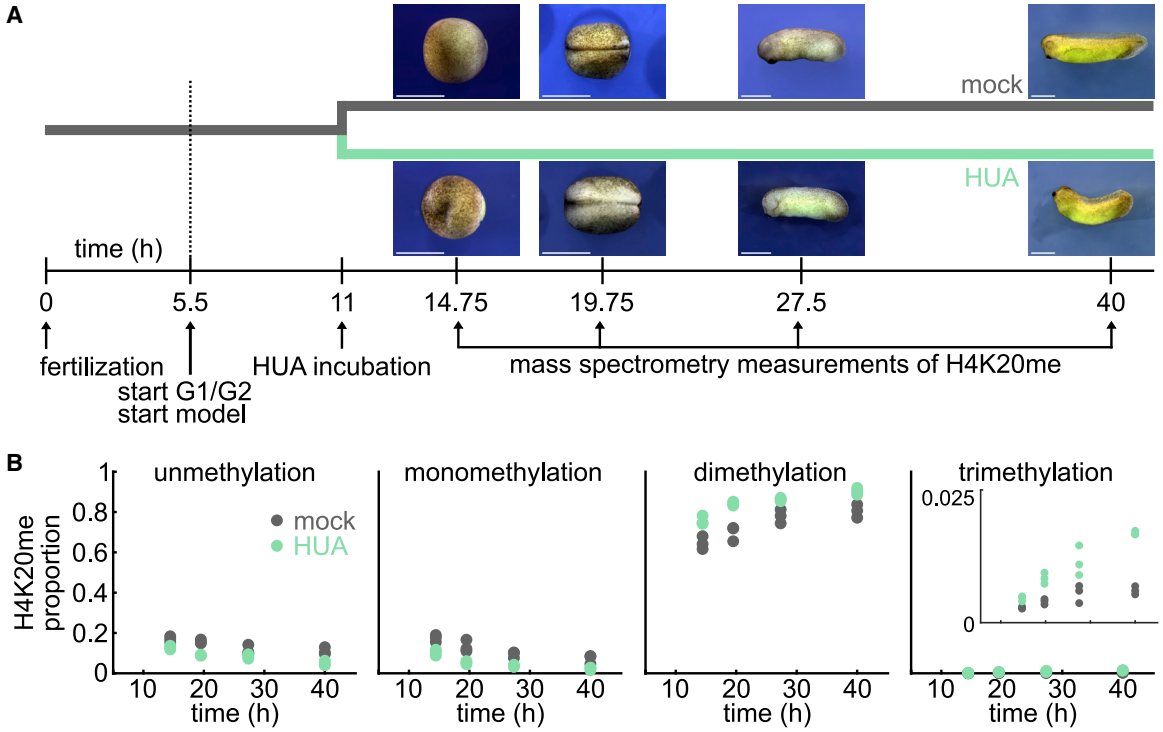


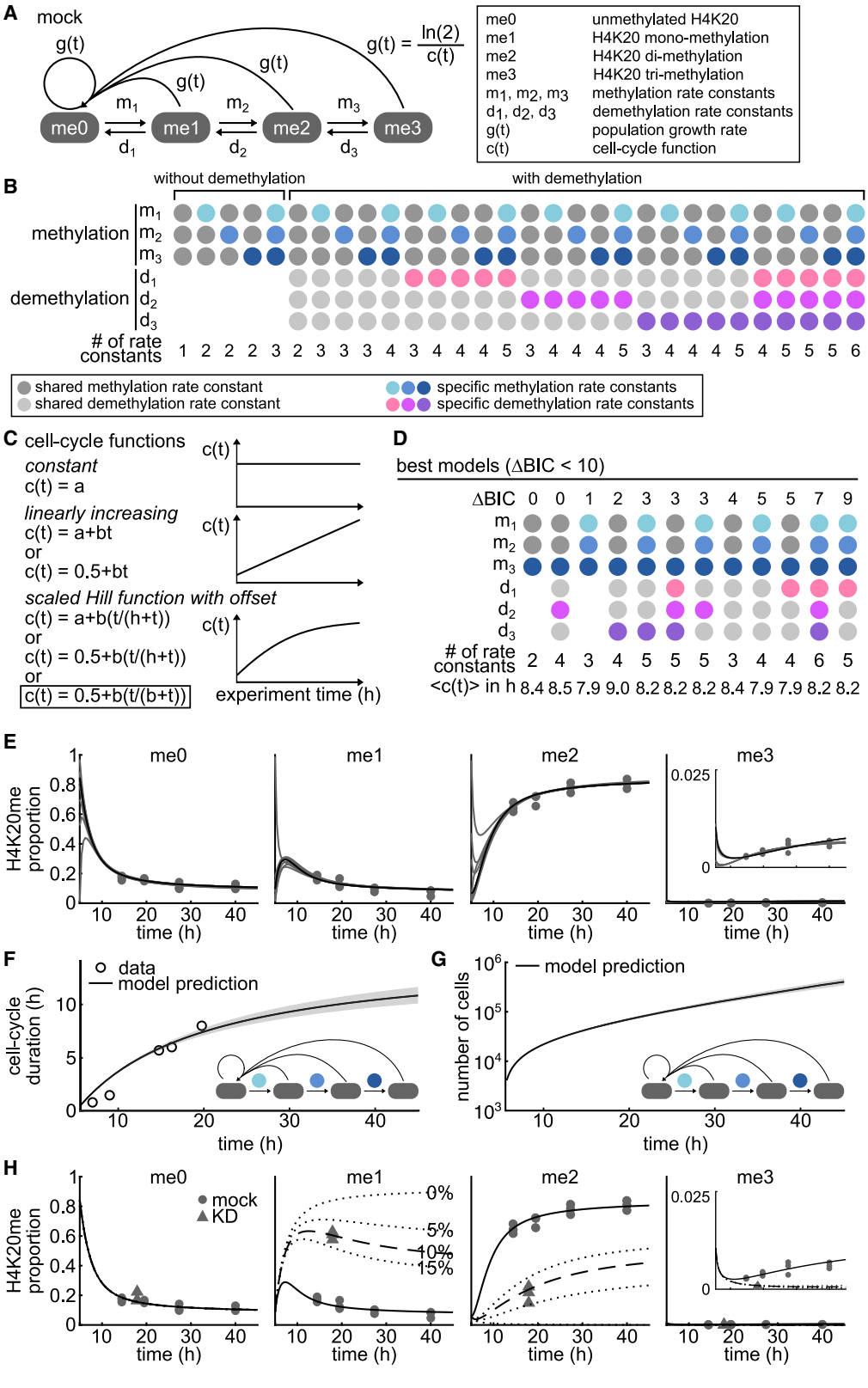
Figure 1. H4K20 Methylation Kinetics during Xenopus Embryogenesis Are Altered upon HUA-Induced Cell-Cycle Arrest

(A) Xenopus eggs are fertilized *in vitro* at time point 0. For the next 5 hpf, the embryonic cell-cycle consists of S and M phases only. At 5.5 hpf, G1 and G2 phases appear. At 11 hpf, half of the embryos are incubated with hydroxyurea/aphidicolin (HUA), arresting cells at the G1/S boundary. Mass spectrometry measurements of H4K20 methylation (H4K20me) are performed at 14.75, 19.75, 27.5, and 40 hpf in embryos with dividing (mock) or non-dividing cells (HUA). HUA incubated embryos are viable and visually remarkably similar to mock embryos (scale bar 1 mm).

(B) H4K20me kinetics differ significantly between mock (gray) and HUA treated (green) embryo populations (two-sample t test for all three biological replicates of mock and HUA for each time point resulted in p values < 0.05 for 15 out of 16 time points). In HUA H4K20 un- and mono-methylation is decreased while H4K20 di- and tri-methylation (see inset) is increased.

constants are derived we will refer to the biological rate constants as rate constants from here on. The most general model contains three rate constants for methylation m_1, m_2, and m_3 and three rate constants for demethylation d_1, d_2, and d_3 (Figure 2B, rightmost model). However, we also considered models with less parameters: rate constants shared between two or more reactions are termed "shared methylation/demethylation rate constants" (Figure 2B, gray) and rate constants specific to one reaction are termed "specific methylation/demethylation rate constants" (Figure 2B, colored). Intrigued by the question whether demethylation is important for methylation kinetics at all (as its existence was recently challenged at least for histone H3 lysine 27 tri-methylation; Reverón-Gómez et al., 2018), we also considered 5 models without demethylation. In total, the 30 models we consider comprise between 1 and 6 rate constants (Figure 2B; STAR Methods). In addition to the rate constants, we inferred another 4 model parameters: 3 initial H4K20me proportions at 5.5 hpf (denoted as $me0_0$, $me1_0$, $me2_0$, $me3_0$ with $me0_0 = 1 - me1_0 - me2_0 - me3_0$), and one noise parameter σ, determining the width of the Laplacian noise distribution (STAR Methods). As we were interested in H4K20me kinetics under the influence of the cell cycle, we started our mock model at 5.5 hpf (Figure 1A), when a regular zygotic cell-cycle with G1/G2 phases is initiated (Newport and Kirschner, 1982). Since cell cycle has been shown to vary substantially with embryonic age, we considered 6 different cell-cycle functions c(t) to model cell cycle over the experiment time t: constant, linearly increasing, or gradually plateauing (using a scaled Hill function with Hill coefficient 1 and offset) (Figure 2C). The number of model parameters for the cell-cycle functions varied from 1 (for the constant cell-cycle function) to 3 parameters (for the gradually plateauing cell-cycle function) (STAR Methods). We performed multi-start maximum likelihood optimization and model selection on 180 models (30 models times 6 different cell-cycle functions). Including prior biological knowledge about the short cell-cycle at 5.5 hpf of ~30 min (Anderson et al., 2017; Gelens et al., 2015), we found that only one of the six tested cell-cycle functions was able to predict a biologically meaningful average cell-cycle duration of around 8 h: a constrained scaled Hill function with Hill coefficient 1 and offset 0.5, $c(t) = 0.5 + b(t/(b + t))$ (Table S1). All models with other cell-cycle functions estimated average cell-cycle durations of at least 70 h. Using a constrained scaled Hill function, we found 12 models that outperformed other models with a BIC (Bayesian information criterion) difference of $\Delta BIC > 10$, which is considered to be an appropriate threshold for model rejection (Kass and Raftery, 1995) (Figure 2D). The two best models (with $\Delta BIC = 0$) show specificity in tri-methylation and shared rate constants for mono- and di-methylation. Overall, the best models with and without demethylation showed specificity

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003



(legend on next page)

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

in either all three methylation rate constants or only in the tri-methylation rate constant. Varying numbers of demethylation rate constants were possible, ranging from 0 to 3. Fits to these 12 top models were able to capture the kinetics underlying H4K20me during mock embryogenesis (Figure 2E). Together, we found that either three specific methylation rate constants or one specific tri-methylation rate constant were necessary to explain the data from untreated embryos and that active demethylation was not required.

Validation of Mock Model by Comparing Cell-Cycle Durations to Experimental Data

We validated one of the best-performing models by comparing it to the average cell-cycle durations experimentally measured in Xenopus neural progenitors at various developmental stages (Graham and Morgan, 1966; Sabherwal et al., 2014; Thuret et al., 2015). We are aware that this comparison is drawn between the average cell-cycle durations of heterogeneous cell populations of Xenopus embryos and potentially more homogeneous cell populations of neural progenitors. However, to the best of our knowledge, this is the only available data on average cell-cycle durations during early Xenopus embryogenesis, which we could use for comparison. The cell-cycle durations from the mock model with three specific methylation rate constants but no demethylation (Figure 2F, inset) showed good agreement with measured cell-cycle durations (Figure 2F). Using this model, we can also predict the absolute number of cells within a normally developing embryo, which is experimentally challenging. For the same model (Figure 2G, inset), the number of cells was predicted to rise exponentially from roughly 20,000 cells after 10 h to 300,000 cells after 40 h (Figure 2G and STAR Methods). Similar results are obtained for the other best-performing mock models. Additionally, we predicted the effect of morpholino knockdowns of the di- and tri-methyltransferases SUV4-20H1/2 (KD) on H4K20me kinetics with the same model (Figure 2H). We found that a complete reduction of the di- and tri-methylation rate constants did not match the data perfectly. However, under the assumption that either the knockdown efficacy is not 100% or that there exist other enzymes performing di- and/or tri-methylation leading to a leaky reduction of the original di- and tri-methylation rate constants to 10%, the model is able to capture the perturbation.

Specific Methylation Rate Constants and Demethylation Are Necessary to Model H4K20me in HUA Embryogenesis

In contrast to mock, methylated H4K20 is not diluted in the cell-cycle-arrested HUA embryo population. We thus modeled HUA with the same set of reactions, however, without a cell-cycle function $g(t) = 0$ (Figure 3A). Similarly, to the mock model we performed multi-start maximum likelihood optimization and model selection on 30 HUA models with and without demethylation. We found that the five best-performing models (with $\Delta BIC < 10$) all required three specific methylation rate constants and demethylation (Figure 3B). The number of demethylation rate constants varied between 0 and 3 (Figure 3B). The single best-performing HUA model without demethylation (rightmost model in Figure 3B) was substantially outperformed by the HUA models with demethylation ($\Delta BIC = 13$), suggesting that demethylation was essential to explain the HUA data. The model fits of the five best HUA models were able to capture the kinetics underlying H4K20me during HUA embryogenesis (Figure 3C). Together, we found that three specific methylation rate constants were necessary to explain the HUA data and that demethylation was essential.

Joint Model Is Able to Retrieve Cell-Cycle Dependence of H4K20me and Finds Demethylation to Be Essential in HUA but Not Necessary in Mock

The models performing best in mock and HUA required three specific methylation rate constants and were indecisive about demethylation ranging from no demethylation over one shared to three

Figure 2. Demethylation Is Not Necessary to Explain Data of Cycling Mock Cells

(A) Model of cycling mock population composed of four H4K20 states: un- (me0), mono- (me1), di- (me2), and tri-methylation (me3). m_1, m_2, and m_3 represent the mono-, di-, and tri-methylation rate constants and d_1, d_2, and d_3 represent the demethylation rate constants. An overall dilution of methylation happens due to cell division, parametrized with population growth rate g(t), which is dependent on the cell-cycle function c(t).

(B) All possible parameter combinations result in 5 models without demethylation and 25 models with demethylation. Rate constants specific to a particular methylation or demethylation step are indicated in color, rate constants shared between methylation or demethylation steps are shown in gray. The number of rate constants ranges between 1 for the simplest model with no demethylation and shared methylation rate constant and 6 for the most complex model, where each methylation and demethylation rate constant is specific.

(C) Only a constrained scaled Hill function with Hill coefficient 1 and offset 0.5 gives an average cell-cycle duration in the expected range of 8 h (marked by the black box). All other cell-cycle functions c(t) predicted average cell-cycle durations of at least 70 h, which is biologically not meaningful and reflects a population of non-cycling cells.

(D) The 12 best-performing models are ordered by increasing BIC. All models with $\Delta BIC < 10$ require either three specific methylation rate constants (m_1, m_2, and m_3) or a specific tri-methylation rate constant. However, if present, demethylation may take on any of the 5 possible rate constant combinations. The best-performing models without demethylation perform similarly well as the best-performing models with demethylation ($\Delta BIC = 0$ and 1). The estimated average cell-cycle duration $<c(t)>$ is in a biologically realistic range of around 8 h.

(E) All 12 best-performing models fit the data. The model with three specific methylation rate constants but with no demethylation is shown in black.

(F) Model prediction of the cell-cycle duration (median, 25^{th} and 75^{th} percentiles of MCMC samples of the cell-cycle parameter of the model with three specific methylation rate constants but with no demethylation (inset)) agrees with experimental measurements of different papers.

(G) The model with three specific methylation rate constants but with no demethylation (inset) predicts an increase of cell numbers from roughly 20,000 cells after 10 h to 300,000 cells after 40 h (using the median, 25^{th} and 75^{th} percentiles of the MCMC samples of the cell-cycle parameter of the model with three specific methylation rate constants but with no demethylation (inset)) in a developing Xenopus embryo.

(H) The model with three specific methylation rate constants but with no demethylation is able to predict the effects on H4K20me upon morpholino knockdowns of the di- and tri-methyltransferases SUV4-20H1/2 (KD) assuming a reduction to 10% of the original di- and tri-methylation rate constants. The dotted lines are the H4K20me kinetics predictions corresponding to 0%, 5%, and 15% of the original di- and tri-methylation rate constants. The solid line shows the previous fit with 100% of the original di-and tri-methylation rate constants.

See also Table S1.

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

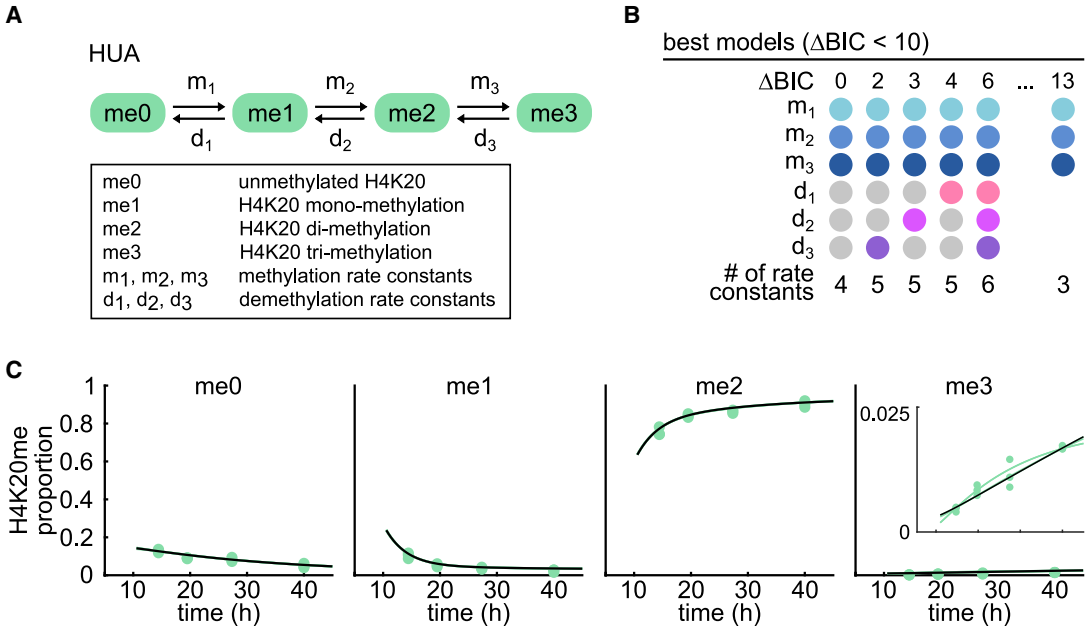


Figure 3. Demethylation Is Essential to Explain Data of Cell-Cycle-Arrested HUA Cells

(A) Model of cell-cycle-arrested HUA population. In contrast to the mock model (Figure 2A), the HUA cells do not divide (g(t) = 0), and no dilution of methylated H4K20 is required.

(B) The 5 best-performing HUA models with ΔBIC < 10 all require 3 specific methylation rate constants (m1, m2, and m3) and demethylation. However, demethylation may take on any of the 5 possible rate constant combinations. The single best-performing HUA model without demethylation (right) is outperformed by the HUA models with demethylation (ΔBIC = 13).

(C) Model fits of top 5 HUA models with demethylation overlap strongly and show the ability to explain the HUA data. The best-performing model is highlighted in black.

See also Table S1.

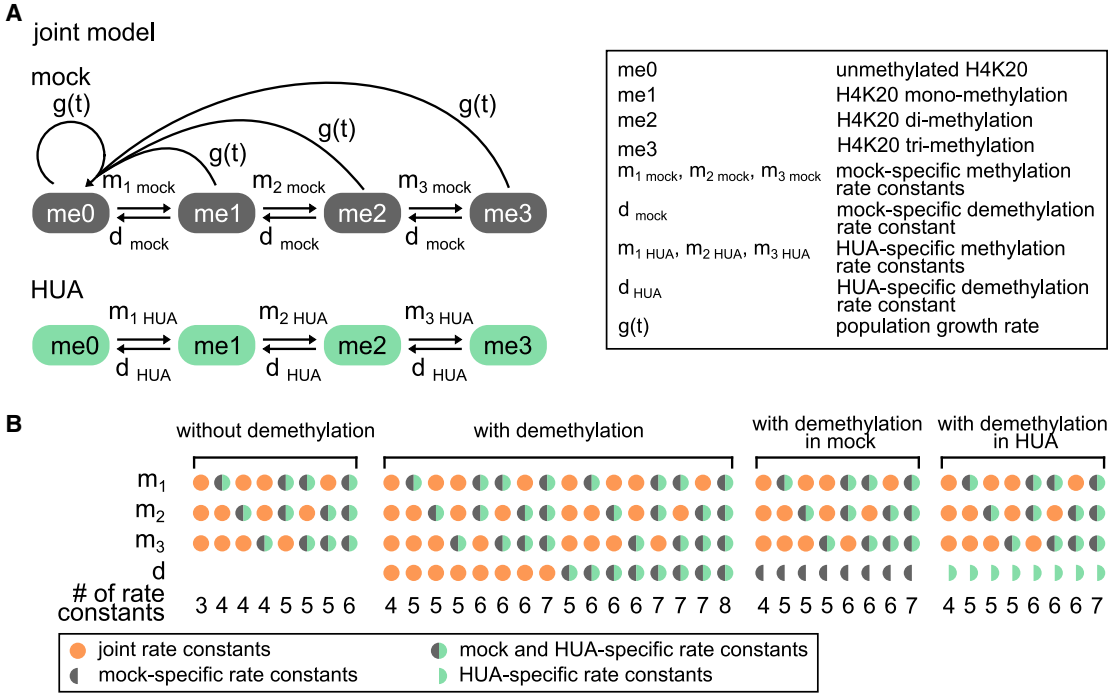
specific demethylation rate constants (Figures 2D and 3B). To determine which rates are substantially different between the two Xenopus populations we considered these findings and devised a joint model considering mock and HUA data simultaneously. For the most general hypothesis (Figure 4A), we allowed for three mock- and three HUA-specific methylation rate constants (visualized by a half gray and half green dot for m_1, m_2, or m_3 in Figure 4B). We also allowed for joint methylation rate constants shared between specific mock and HUA methylation steps (visualized as an orange dot for m_1, m_2, or m_3 in Figure 4B) reducing the number of parameters. As demethylation was not necessary to explain the mock data and one demethylation rate constant shared between methylation steps was sufficient for HUA, we here restricted demethylation to the simplest case of at most one shared demethylation rate constant d_{mock} and d_{HUA} (Figure 4A). We allowed for mock- and HUA-specific demethylation rate constants (visualized again by a half gray and half green dot for d in Figure 4B) or a joint demethylation rate constant for mock and HUA (visualized as an orange dot for d in Figure 4B). Furthermore, as a constrained scaled Hill function with Hill coefficient 1 and offset 0.5 was the only function that led to biologically meaningful cell-cycle durations (see above and Figures 2C and 2F), we did not consider different cell-cycle functions thereby reducing the set of possible models to 8 joint models without demethylation (Figure 4B left), 16 joint models with demethylation and 2 × 8 models with demethylation in either mock or HUA (Figure 4B right). To identify joint models that are able to explain our measured data, we again fitted the models using multi-start maximum likelihood optimization and model selection.

All 6 best-performing joint models (ΔBIC < 10) required mock- and HUA-specific mono- and di-methylation rate constants (Figure 4C). However, they were not conclusive about tri-methylation and, if present, demethylation rate constants (Figure 4C). Specificity in one or more rate constants highlights that the differences in H4K20me proportions of mock and HUA are not explicable by the missing cell cycle alone but that the overall H4K20me kinetics are cell-cycle dependent. The model structure of the best-performing joint model (model I) is shown in Figure 4D. Joint models with demethylation in HUA only (models I and V in Figure 4C) performed just as well as joint models with demethylation in both HUA and mock (models II, III, IV, and VI in Figure 4C) while joint models without demethylation (ΔBIC = 17 and 20) and joint models with demethylation in mock only (ΔBIC = 21 and 26) were substantially outperformed. This suggested that demethylation was essential for HUA only, in accordance with the results from the separate models (Figures 2D and 3B).

The top 6 joint models (models I–VI in Figure 4C) showed good overall agreement with mock and HUA data (Figure 4E) and strongly consistent rate constants (Figure 4F). We determined the marginal distributions for all rate constants by Markov chain Monte Carlo (MCMC) sampling, where the credibility ranges are the 25th and 75th percentiles of the marginal distributions (see STAR Methods). We found strong discrepancies between mono- and di-methylation rate constants for mock and HUA, decreasing

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

10-fold and 2-fold, respectively (Figure 4F). The mock-specific mono- and di-methylation rate constants of the top 6 joint models had overlapping credibility ranges suggesting that for mock, a shared rate constant for mono- and di-methylation would suffice (Figure 4F). Similarly, mock- and HUA-specific tri-methylation rate constants show overlapping credibility ranges suggesting that a joint tri-methylation rate constant would suffice. In joint models with demethylation (models III and VI) we found the mock-specific demethylation rate constants to take on very small values while the HUA-specific demethylation rate constants were small but substantially larger: for model VI the median mock-specific demethylation rate constant was estimated to be $2.0\ 10^{-4}$ (with $0.5–8.3\ 10^{-4}$ credibility range), while the HUA-specific demethylation rate constant was estimated to be $5.9\ 10^{-3}$ ($4.9–7.0\ 10^{-3}$) (Figure 4F).

DISCUSSION

The joint demethylation rate constants (in models II and IV) were estimated to similar values as HUA-specific demethylation rate constants (the joint demethylation rate constant for model II and the HUA-specific demethylation rate constant for model VI were both estimated to be $5.9\ 10^{-3}$ [$4.9–7.0\ 10^{-3}$]). This suggests that joint demethylation rate constants are overshadowed by the HUA model, strengthening the hypothesis that demethylation is not necessary in mock but essential in HUA. We would like to note that we here consider bulk H4K20me mass spectrometry data. However, demethylation might act highly localized and specific on only a few promoter nucleosomes with very important functions. Biologically, this could mean that while demethylases are present during embryogenesis, their effect in cycling cells is minute due to an overall dilution by unmodified histones. Only when cells stop to cycle (as modeled with the HUA treatment in our approach) demethylation kicks in and stabilizes post-translational modifications specifically, thereby potentially driving differentiation. To verify this experimentally, a knockdown of the H4K20 demethylases in HUA-treated embryos would be required where we hypothesize severe phenotypes due to the cell's incapability of reversing methylation. In contrast, such a knockdown should show little to no effect in untreated embryos as we hypothesize little to no active demethylation here. However, known H4K20 demethyltransferases e.g., PHF8, ROSBIN, and PHF2 are not specific to H4K20, and no global inhibitors of H4K20 demethylation are yet known. Hence, an experimental validation of our model predictions is currently not feasible due to technical limitations and therefore beyond the scope of this work.

Our findings can be interpreted in light of the current knowledge on methyltransferases. The mono-methyltransferase KMT5A (also known as PR-Set7) was found to be cell-cycle dependent, getting degraded by the proteasome in G1 phase (Abbas et al., 2010; Centore et al., 2010). In the absence of KMT5A, mono-methylation might be compensated by SUV4-20H1/2 but with lower activity (Southall et al., 2014; Yang et al., 2008). HUA treatment blocks the cell-cycle at the G1/S boundary, suggesting that none to little KMT5A is present in HUA to mono-methylate H4K20. This is reflected by a 10-fold decrease in the HUA-specific mono-methylation rate constant in all best-performing joint models (Figure 4F). As HUA-specific mono-methylation rate constants were necessary to explain the data (see Figure 4C), the joint model is able to retrieve this known cell-cycle dependence of H4K20me. H4K20me2 is also regulated in a cell-cycle dependent manner, however, peaking in G1 phase (Pesavento et al., 2008). In contrast, all best-performing joint models estimate the HUA-specific di-methylation rate constants to be decreased 2-fold in comparison to the mock-specific di-methylation (Figure 4F). We hypothesize this unexpected decrease of HUA-specific di-methylation to be due to either compensatory effects of SUV4-20H1/2, when the enzymes additionally mono-methylate H4K20, or so far unknown effects.

The separate model for mock identified either only tri-methylation or all three methylation steps to be specific (Figure 2D). The joint model reflects the same specificities regarding methylation in mock. Even though the joint model allows for specificity in all three methylation steps, the credibility ranges of mock-specific mono- and di-methylation rate constants in the joint models overlap (Figure 4F left). This suggests that a shared rate constant for mock mono- and di-methylation would suffice resulting in a mock model with specificity in tri-methylation only. However, in the joint models the HUA-specific mono- and di-methylation rate constants have non-overlapping credibility ranges with respect to the mock-specific rate constants (10-fold and 2-fold decrease) nor to each other. Under the assumption that the mock and HUA models are based on the same model structure, allowing for three

Figure 4. Joint Computational Modeling Allows Direct Comparisons between Mock and HUA Rate Constants and Reveals that Demethylation Is Overshadowed by HUA

(A) Joint model allows for three methylation and one demethylation rate constants for both mock and HUA as suggested by the best models for mock and HUA.

(B) We fit 16 models with demethylation and 8 models each for without demethylation in mock and/or HUA to the joint data to infer mock- and HUA-specific rate constants. The joint rate constants of mock and HUA are shown in orange, the rate constants present in both the mock and HUA models but taking on mock- and HUA-specific values are indicated in gray/green, the rate constants only present in the mock or HUA model are shown in gray and green half-circles, respectively. The model structure of the most complex of models is shown in (A). The number of rate constants ranges between 3 and 8.

(C) The best-performing models on the combined dataset are ordered according to their BIC value. All models require HUA-specific mono- and di-methylation rate constants but are indecisive about tri- methylation and demethylation. Joint models where demethylation is present in either only HUA or in both mock and HUA perform equally well. Joint models where demethylation is not present in either only HUA or in both mock and HUA perform considerably worse. Model IDs of all considerably best-performing models are given (I–VI).

(D) Model structure of the simplest best-performing joint model with demethylation in only HUA (model I).

(E) All best-performing joint models are able to explain both the mock and HUA data. The estimated initial conditions vary between the models. Joint model I is highlighted.

(F) The violin plots of the marginal distributions of all best-performing joint models show high consistency between the estimated methylation and demethylation rate constants. HUA-specific mono- and di-methylation rate constants are considerably decreased. Tri-methylation rate constants between mock and HUA have strongly overlapping marginal distributions. Demethylation seems to be dominated by the HUA population and is negligible in the mock population if a mock-specific demethylation rate is allowed.

See also Table S1.

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

specific methylation rate constants in the joint model was thus necessary (Figure 4F) to resolve these differences.

All three joint models with specificity in tri-methylation (models IV, V, and VI) result in slightly lower BIC values (Figure 4C), which is likely due to an increased penalization term for an additional estimated parameter and not due to a decreased likelihood. The estimated tri-methylation rate constants are small (on the order of 10^{-3}) and the credibility ranges for mock- and HUA-specific tri-methylation overlap in all three joint models suggesting that a joint tri-methylation rate constant would suffice. When we interpret differences in HUA and mock rates as indications for cell-cycle dependent rates, we find no evidence for cell-cycle dependence for H4K20 tri-methylation. To clarify if the corresponding enzymes are indeed homogeneously expressed is up to further research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILIBILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Embryos Handling and HUA Treatment
  - Nuclear Histone Extraction
  - Mass Spectrometry Sample Preparation
  - Mass Spectrometry Analysis
  - Histone Modifications Quantification
- METHOD DETAILS
  - Models
  - Noise Models
- OPTIMIZATION AND PARAMETER ESTIMATION
  - Model Selection
  - Parameter Uncertainty
  - Validation - Cell-Cycle Durations
  - Prediction of Number of Cells
  - Implementation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.11.003.

ACKNOWLEDGMENTS

L.S. was funded by the BMBF project TIDY (031L0170B). L.S. is especially grateful to the Technical University of Munich's Department of Mathematics, whose generous Entrepreneurial Award (within the Program "Global Challenges for Women in Math Science") contributed to the completion of this project. D.P., A.I., and R.R. were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project ID213249687 – SFB 1064. C.M. acknowledges funding from the European Research Council (ERC), grant agreement no. 866411.

AUTHOR CONTRIBUTIONS

Conceptualization, A.I., R.A.W.R., and C.M.; Methodology, L.S. and C.L.; Software, L.S. and C.L. Validation, C.L.; Formal Analysis, L.S.; Investigation, L.S. and D.P.; Resources, A.I., R.A.W.R., and C.M.; Data Curation, D.P.; Writing – Original Draft, L.S.; Visualization, L.S.; Supervision, A.I., R.A.W.R., C.L., and C.M.; Project Administration, A.I., R.A.W.R., and C.M.; Funding Acquisition, A.I., R.A.W.R., and C.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 14, 2020
Revised: August 5, 2020
Accepted: November 11, 2020
Published: December 8, 2020

REFERENCES

Abbas, T., Shibata, E., Park, J., Jha, S., Karnani, N., and Dutta, A. (2010). CRL4(Cdt2) regulates cell proliferation and histone gene expression by targeting PR-Set7/Set8 for degradation. Mol. Cell *40*, 9–21.

Alabert, C., Loos, C., Voelker-Albert, M., Graziano, S., Forné, I., Reveron-Gomez, N., Schuh, L., Hasenauer, J., Marr, C., Imhof, A., and Groth, A. (2020). Domain model explains propagation dynamics and stability of histone H3K27 and H3K36 methylation landscapes. Cell Rep. *30*, 1223–1234.e8.

Anderson, G.A., Gelens, L., Baker, J.C., and Ferrell, J.E., Jr. (2017). Desynchronizing embryonic cell division waves reveals the robustness of Xenopus laevis development. Cell Rep. *21*, 37–46.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Boterenbrood, E.C., Narraway, J.M., and Hara, K. (1983). Duration of cleavage cycles and asymmetry in the direction of cleavage waves prior to gastrulation inXenopus laevis. Wilhelm Rouxs Arch. Dev. Biol. *192*, 216–221.

Bowes, J.B., Snyder, K.A., Segerdell, E., Jarabek, C.J., Azam, K., Zorn, A.M., and Vize, P.D. (2010). Xenbase: gene expression and improved integration. Nucleic Acids Res. *38*, D607–D612.

Centore, R.C., Havens, C.G., Manning, A.L., Li, J.M., Flynn, R.L., Tse, A., Jin, J., Dyson, N.J., Walter, J.C., and Zou, L. (2010). CRL4(Cdt2)-mediated destruction of the histone methyltransferase Set8 prevents premature chromatin compaction in S phase. Mol. Cell *40*, 22–33.

Evertts, A.G., Manning, A.L., Wang, X., Dyson, N.J., Garcia, B.A., and Coller, H.A. (2013). H4K20 methylation regulates quiescence and chromatin compaction. Mol. Biol. Cell *24*, 3025–3037.

Faber, J., and Nieuwkoop, P.D. (1994). Normal Table of Xenopus Laevis (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg Till the End of Metamorphosis (Garland Science).

Fang, J., Feng, Q., Ketel, C.S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J.A., and Zhang, Y. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. Curr. Biol. *12*, 1086–1099.

Feng, W., Yonezawa, M., Ye, J., Jenuwein, T., and Grummt, I. (2010). PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. Nat. Struct. Mol. Biol. *17*, 445–450.

Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat. Genet. *37*, 391–400.

Fröhlich, F., Kaltenbacher, B., Theis, F.J., and Hasenauer, J. (2017). Scalable parameter estimation for genome-scale biochemical reaction networks. PLoS Comput. Biol. *13*, e1005331.

Gelens, L., Huang, K.C., and Ferrell, J.E., Jr. (2015). How does the Xenopus laevis embryonic cell cycle avoid spatial chaos? Cell Rep. *12*, 892–900.

Graham, C.F., and Morgan, R.W. (1966). Changes in the cell cycle during early amphibian development. Dev. Biol. *14*, 439–460.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. *13*, 343–357.

Harris, W.A., and Hartenstein, V. (1991). Neuronal determination without cell division in Xenopus embryos. Neuron *6*, 499–515.

Hass, H., Loos, C., Raimúndez-Álvarez, E., Timmer, J., Hasenauer, J., and Kreutz, C. (2019). Benchmark problems for dynamic modeling of intracellular processes. Bioinformatics *35*, 3073–3082.

Heasman, J. (2006). Patterning the early Xenopus embryo. Development *133*, 1205–1217.

Jadhav, U., Manieri, E., Nalapareddy, K., Madha, S., Chakrabarti, S., Wucherpfennig, K., Barefoot, M., and Shivdasani, R.A. (2020). Replicational dilution of H3K27me3 in mammalian cells and the role of poised promoters. Mol. Cell *78*, 141–151.e5.

Jambhekar, A., Dhall, A., and Shi, Y. (2020). Author Correction: roles and regulation of histone methylation in animal development. Nat. Rev. Mol. Cell Biol. *21*, 59.

Jasencakova, Z., Scharf, A.N.D., Ask, K., Corpet, A., Imhof, A., Almouzni, G., and Groth, A. (2010). Replication stress interferes with histone recycling and predeposition marking of new histones. Mol. Cell *37*, 736–743.

Jørgensen, S., Schotta, G., and Sørensen, C.S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. Nucleic Acids Res *41*, 2797–2806.

Kass, R.E., and Raftery, A.E. (1995). Bayes factors. J. Am. Stat. Assoc. *90*, 773–795.

Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA methylation in cancer and aging. Cancer Res. *76*, 3446–3450.

Lachner, M., Sengupta, R., Schotta, G., and Jenuwein, T. (2004). Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. Cold Spring Harbor Symp. Quant. Biol. *69*, 209–218.

Leroy, G., Dimaggio, P.A., Chan, E.Y., Zee, B.M., Blanco, M.A., Bryant, B., Flaniken, I.Z., Liu, S., Kang, Y., Trojer, P., and Garcia, B.A. (2013). A quantitative atlas of histone modification signatures from human cancer cells. Epigenet. Chromatin *6*, 20.

Liebler, D.C., and Zimmerman, L.J. (2013). Targeted Quantitation of Proteins by Mass Spectrometry. Biochemistry *52*, 3797–3806.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968.

Maier, C., Loos, C., and Hasenauer, J. (2017). Robust parameter estimation for dynamical systems from outlier-corrupted data. Bioinformatics *33*, 718–725.

Newport, J., and Kirschner, M. (1982). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675–686.

Nicetto, D., Hahn, M., Jung, J., Schneider, T.D., Straub, T., David, R., Schotta, G., and Rupp, R.A.W. (2013). Suv4-20h histone methyltransferases promote neuroectodermal differentiation by silencing the pluripotency-associated Oct-25 gene. PLoS Genet. *9*, e1003188.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., et al. (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol. Cell *9*, 1201–1213.

Pesavento, J.J., Yang, H., Kelleher, N.L., and Mizzen, C.A. (2008). Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol. Cell. Biol. *28*, 468–486.

Pokrovsky, D., Forne, I., Straub, T., Imhof, A., and Rupp, R. (2020). Mitotic activity shapes stage-specific histone modification profiles during Xenopus embryogenesis. bioRxiv https://www.biorxiv.org/content/10.1101/2020.08.04.200550v1?rss=1.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Reverón-Gómez, N., González-Aguilera, C., Stewart-Morgan, K.R., Petryk, N., Flury, V., Graziano, S., Johansen, J.V., Jakobsen, J.S., Alabert, C., and Groth, A. (2018). Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication. Mol. Cell *72*, 239–249.e5.

Sabherwal, N., Thuret, R., Lea, R., Stanley, P., and Papalopulu, N. (2014). aPKC phosphorylates p27Xic1, providing a mechanistic link between apicobasal polarity and cell-cycle control. Dev. Cell *31*, 559–571.

Sakaguchi, A., and Steward, R. (2007). Aberrant monomethylation of histone H4 lysine 20 activates the DNA damage checkpoint in Drosophila melanogaster. J. Cell Biol. *176*, 155–162.

Sanders, S.L., Portoso, M., Mata, J., Bähler, J., Allshire, R.C., and Kouzarides, T. (2004). Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. Cell *119*, 603–614.

Schneider, T.D., Arteaga-Salas, J.M., Mentele, E., David, R., Nicetto, D., Imhof, A., and Rupp, R.A.W. (2011). Stage-specific histone modification profiles reveal global transitions in the Xenopus embryonic epigenome. PLoS One *6*, e22548.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. *18*, 1251–1262.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callén, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., et al. (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev. *22*, 2048–2061.

Schwarz, G. (1978). Estimating the dimension of a model. Ann. Statist. *6*, 461–464.

Sive, H.L., Grainger, R.M., and Harland, R.M. (2000). Early development of Xenopus laevis: a laboratory manual (Vancouver: CSHL Press).

Southall, S.M., Cronin, N.B., and Wilson, J.R. (2014). A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. Nucleic Acids Res. *42*, 661–671.

Stapor, P., Weindl, D., Ballnus, B., Hug, S., Loos, C., Fiedler, A., Krause, S., Hroß, S., Fröhlich, F., and Hasenauer, J. (2018). PESTO: parameter EStimation TOolbox. Bioinformatics *34*, 705–707.

Thuret, R., Auger, H., and Papalopulu, N. (2015). Analysis of neural progenitors from embryogenesis to juvenile adult in Xenopus laevis reveals biphasic neurogenesis and continuous lengthening of the cell cycle. Biol. Open *4*, 1772–1781.

van Kruijsbergen, I., Hontelez, S., Elurbe, D.M., van Heeringen, S.J., Huynen, M.A., and Veenstra, G.J.C. (2017). Heterochromatic histone modifications at transposons in Xenopus tropicalis embryos. Dev. Biol. *426*, 460–471.

van Nuland, R., and Gozani, O. (2016). Histone H4 lysine 20 (H4K20) methylation, expanding the signaling potential of the proteome one methyl moiety at a time. Mol. Cell. Proteomics *15*, 755–764.

Villar-Garea, A., Forne, I., Vetter, I., Kremmer, E., Thomae, A., and Imhof, A. (2012). Developmental regulation of N-terminal H2B methylation in Drosophila melanogaster. Nucleic Acids Res. *40*, 1536–1549.

Xiao, B., Jing, C., Kelly, G., Walker, P.A., Muskett, F.W., Frenkiel, T.A., Martin, S.R., Sarma, K., Reinberg, D., Gamblin, S.J., and Wilson, J.R. (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes Dev. *19*, 1444–1454.

Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. J. Biol. Chem. *283*, 12085–12092.

Zee, B.M., Britton, L.M., Wolle, D., Haberman, D.M., and Garcia, B.A. (2012). Origins and formation of histone methylation across the human cell cycle. Mol. Cell. Biol. *32*, 2503–2514.

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
H4K20me proportions	This work	https://doi.org/10.5281/zenodo.4046502
Software and Algorithms		
MATLAB2017a (including the Statistics and Optimization Toolbox)	Mathworks	https://www.mathworks.com
AMICI	Fröhlich et al., 2017	http://icb-dcm.github.io/AMICI/
PESTO	Stapor et al., 2018	https://github.com/ICB-DCM/PESTO
Code – parameter estimation and model selection	This work	https://doi.org/10.5281/zenodo.4046502

RESOURCE AVAILIBILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Carsten Marr (carsten.marr@helmholtz-muenchen.de).

Materials Availability

This study did not generate new materials.

Data and Code Availability

- H4K20 methylation proportions have been deposited at https://github.com/marrlab/HistonesXenopus and are publicly available at https://doi.org/10.5281/zenodo.4046502.
- Original MATLAB code is publicly available at https://github.com/marrlab/HistonesXenopus and https://doi.org/10.5281/zenodo.4046502.
- The scripts used to generate the figures reported in this paper are available at https://github.com/marrlab/HistonesXenopus and https://doi.org/10.5281/zenodo.4046502.
- Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal work has been conducted in accordance with Deutsches Tierschutzgesetz; Xenopus experiments were approved by the Government of Oberbayern.

Embryos Handling and HUA Treatment

Xenopus laevis eggs were collected, in vitro fertilized and handled by standard methods (Sive et al., 2000). The staging was done according to Nieukoop and Faber (Faber and Nieuwkoop, 1994). When embryos reached the desired stage (NF10.5), they were separated into two groups and incubated continuously into either HUA or mock solutions in parallel. HUA solution: 20mM Hydroxyurea (USBiological, H9120) and 150μM Aphidicolin (BioViotica, BVT-0307) in 0.1x MBS solution (Harris and Hartenstein, 1991). Mock solution: 2% DMSO (dissolvent for Aphidicolin) in 0.1x MBS solution. The embryos were collected at the four developmental stages (NF13, NF18, NF25 and NF32) for the mass spectrometry analysis.

Nuclear Histone Extraction

Around 50 to 200 embryos developed to desired stages (NF13, NF18, NF25 NF32). They were harvested and histone proteins were purified by acid extraction from nuclei (Pokrovsky et al., 2020; Schneider et al., 2011). Each developmental stage is represented by three biological replicates. Each biological replicate derived from a different mating pair.

Mass Spectrometry Sample Preparation

The pellet from the nuclear histone extraction was dissolved in an appropriate amount of Lämmli Buffer to reach 1.37 10^6 nuclei/μl in each sample. 15μL were loaded on an 8-16% gradient SDS-PAGE gel (SERVA Lot V140115-1) and stained with Coomassie Blue to

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

visualize the histone bands. Histone bands were excised and propionylated (as described in (Villar-Garea et al., 2012)). As an internal and inter-sample control, a library consisting of heavy-labelled peptides mimicking H4K20 methylation states which contain a heavy Arginine (R10 peptides) was used (product of JPT company). R10 peptides were mixed in the library with the equimolar concentration and the mix was added to each analyzed sample before in-gel trypsin digestion. Digested peptides were sequentially desalted using C18 Stagetips (3M Empore) and porous carbon material (TipTop Carbon, Glygen) as described in (Rappsilber et al., 2007) and resuspended in 15μl of 0.1% FA.

Mass Spectrometry Analysis

To identify and measure the proportion of the histone modifications a parallel reaction monitoring method (PRM) was used (Liebler and Zimmerman, 2013). The mass spectrometer was operated in the scheduled PRM mode to identify and quantify specific fragment ions of N-terminal peptides histone proteins. In this mode, the mass spectrometer automatically switched between one survey scan and 9 MS/MS acquisitions of the m/z values described in the inclusion list containing the precursor ions, modifications and fragmentation conditions. Survey full scan MS spectra (from m/z 270–730) were acquired with resolution 60,000 at m/z 400 (AGC target of 3x10^6). PRM spectra were acquired with resolution 30,000 to a target value of 2 10^5, maximum IT 60 ms, isolation window 0.7 m/z and fragmented at 27% or 30% normalized collision energy. Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C.

Histone Modifications Quantification

Data analysis was performed with Skyline (version 3.7) (MacLean et al., 2010) by using doubly and triply charged peptide masses for extracted ion chromatograms (XICs). Selection of respective peaks was identified based on the retention time and fragmentation spectra of the spiked in heavy-labelled peptides. Integrated peak values (Total Area MS1) were exported as csv file for further calculations. Total area MS1 from endogenous peptides was normalized to the respective area of heavy-labelled peptides. The sum of all normalized total area MS1 values of the same isotopically modified peptide in one sample resembled the amount of total peptide. The proportions of the different K20 methylation states were calculated and displayed as percentages of the overall K20 peptide amount.

METHOD DETAILS

Models

HUA and Mock Models

We consider the proportions of un- (me0), mono- (me1), di- (me2) and tri-methylated (me3) H4K20 within a Xenopus embryo population, defined as

$$\mathrm{meX} = \frac{\mathrm{meX_{MS}}}{\sum_{i=0}^{3} \mathrm{mei_{MS}}},$$

where $\mathrm{meX_{MS}}$ is the H4K20 methylation as measured by mass spectrometry and $X \in \{0,1,2,3\}$. We assume successive methylation and demethylation of H4K20 (van Nuland and Gozani, 2016) resulting in three possible methylation rate constants for mono-, di-, and tri-methylation with rate constants m_1, m_2, m_3, respectively, and three possible demethylation rates with rate constants d_1, d_2, d_3 (Figures 2A and 3A). However, reactions might share rate constants. The simplest model (Figure 2B left) comprises one shared methylation rate constant for mono-, di- and tri-methylation We successively added model-specific rate constants to this simplest model (Figure 2B). Models allowing for two specific methylation rate constants are identical to a model allowing for three specific methylation rate constants. Hence, we do not consider models with two specific methylation rate constants separately. This results in 2^3-3 = 5 models for methylation - three methylation rate constants with either a shared or specific rate constant minus the three cases where we assume only two of the three rate constants to be specific. We have the same for demethylation resulting overall in (2^3-3) × (2^3-3) = 25 possible HUA models.

Joint Models

The joint model considers both mock and HUA data sets. We based the joint model on our previous findings assuming three specific methylation rate constants and at most one demethylation rate constant for both mock (Figure 2D) and HUA (Figure 3B) as well as a scaled Hill function with Hill coefficient 1 and offset 0.5 as cell-cycle function. In general, the joint model would allow for (2^3-3)^4 = 625 distinct models. By constraining both the HUA and mock model to allow for three methylation and one demethylation rate constants, we are able to reduce the number of possible models to 16. The simplest joint model is comprised of 3 rate constants which are shared for mock and the HUA reactions (Figure 4B left). We successively added model complexity by allowing for HUA-specific rate constants, totaling to 16 models for the joint model with demethylation in mock and HUA and 8 models for the joint model with demethylation present in either one or none (Figure 4B).

For all models we describe the temporal changes in these proportions by systems of ordinary differential equations (ODEs) using mass action kinetics (see below).

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

HUA Model

We first derive the system of ODEs for the absolute numbers of H4K20me states, given by mẽ0, mẽ1, mẽ2, and mẽ3, for the model with the largest number of rate constants (Figure 3A):

$$\dot{\tilde{me}}0 = -m_1 \times \tilde{me}0 + d_1 \times \tilde{me}1$$

$$\dot{\tilde{me}}1 = m_1 \times \tilde{me}0 - (m_2 + d_1) \times \tilde{me}1 + d_2 \times \tilde{me}2$$

$$\dot{\tilde{me}}2 = m_2 \times \tilde{me}1 - (m_3 + d_2) \times \tilde{me}2 + d_3 \times \tilde{me}3$$

$$\dot{\tilde{me}}3 = m_3 \times \tilde{me}2 - d_3 \times \tilde{me}3$$

$$\dot{N} = 0,$$

where N is the total number of histone tails. As the HUA model assumes no cell-cycle, the number of histones over time is constant and its derivative is zero. The proportions me0, me1, me2 and me3 are given by $meX = \frac{\tilde{meX}}{N}$, for $X \in \{0, 1, 2, 3\}$ (Alabert et al., 2020) and the corresponding ODEs are given by

$$\dot{meX} = \frac{\dot{\tilde{meX}}}{N} - \frac{\tilde{meX} \times \dot{N}}{N^2}$$

simplifying to

$$\dot{meX} = \frac{\dot{\tilde{meX}}}{N}$$

in the HUA model. The full ODE system for the proportions is given by

$$\begin{aligned}
\dot{me}0 &= -m_1 \times me0 + d_1 \times me1 \\
\dot{me}1 &= m_1 \times me0 - (m_2 + d_1) \times me1 + d_2 \times me2 \\
\dot{me}2 &= m_2 \times me1 - (m_3 + d_2) \times me2 + d_3 \times me3 \\
\dot{me}3 &= m_3 \times me2 - d_3 \times me3 \\
\dot{N} &= 0.
\end{aligned}$$

Mock Model - Constant Cell-Cycle Duration

According to the HUA model, we first formulate the ODE system of the absolute numbers of methylation states, mẽ0, mẽ1, mẽ2, and mẽ3. During DNA replication newly synthesized and unmodified histones are incorporated, leading to a constant increase in unmethylated H4K20 with a population growth rate g(t)=ln(2)/c(t), where c(t) is the cell-cycle function that allows cell-cycle durations to change with time. In the simplest case, we assume the cell-cycle duration to be constant over time, denoted by a:

$$c(t) = a.$$

Then the full ODE system of the absolute numbers of methylation states is given by

$$\dot{\tilde{me}}0 = -m_1 \times \tilde{me}0 + d_1 \times \tilde{me}1 + \frac{\ln(2)}{a} \times (\tilde{me}0 + \tilde{me}1 + \tilde{me}2 + \tilde{me}3)$$

$$\dot{\tilde{me}}1 = m_1 \times \tilde{me}0 - (m_2 + d_1) \times \tilde{me}1 + d_2 \times \tilde{me}2$$

$$\dot{\tilde{me}}2 = m_2 \times \tilde{me}1 - (m_3 + d_2) \times \tilde{me}2 + d_3 \times \tilde{me}3$$

$$\dot{\tilde{me}}3 = m_3 \times \tilde{me}2 - d_3 \times \tilde{me}3$$

$$\dot{N} = \frac{\ln(2)}{a} \times N,$$

where N is the total number of histone tails, $N(t) = N0 \times e^{\frac{\ln(2)}{a} \times t}$ and $N(t_0) = N0$ the number of histone tails at the beginning of the model. Then the ODE system of the methylation proportions is given by

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

$$\dot{me0} = \frac{-m_1 \times \tilde{me0} + d_1 \times \tilde{me1} + \frac{\ln(2)}{a} \times (\tilde{me0} + \tilde{me1} + \tilde{me2} + \tilde{me3})}{N} - \frac{\tilde{me0} \times \frac{\ln(2)}{a} \times N}{N^2}$$

$$= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a} \times (me0 + me1 + me2 + me3) - me0 \times \frac{\ln(2)}{a}$$

$$= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a} \times (me1 + me2 + me3)$$

$$\dot{me1} = m_1 \times me0 - \left(m_2 + d_1 + \frac{\ln(2)}{a}\right) \times me1 + d_2 \times me2$$

$$\dot{me2} = m_2 \times me1 - \left(m_3 + d_2 + \frac{\ln(2)}{a}\right) \times me2 + d_3 \times me3$$

$$\dot{me3} = m_3 \times me2 - \left(d_3 + \frac{\ln(2)}{a}\right) \times me3$$

$$\dot{N} = \frac{\ln(2)}{a} \times N,$$

leading to a constant increase of the unmethylated H4K20 proportion and a constant decrease of the methylated H4K20me proportions.

Mock Model - Linearly Increasing Cell-Cycle

In the case of a linear cell-cycle function

$$c(t) = a + b \times t$$

we first derive the ODE system for the absolute numbers of H4K20 methylation mẽ0, mẽ1, mẽ2, and mẽ3:

$$\dot{\tilde{me0}} = -m_1 \times \tilde{me0} + d_1 \times \tilde{me1} + \frac{\ln(2)}{a + b \times t} \times (\tilde{me0} + \tilde{me1} + \tilde{me2} + \tilde{me3})$$

$$\dot{\tilde{me1}} = m_1 \times \tilde{me0} - (m_2 + d_1) \times \tilde{me1} + d_2 \times \tilde{me2}$$

$$\dot{\tilde{me2}} = m_2 \times \tilde{me1} - (m_3 + d_2) \times \tilde{me2} + d_3 \times \tilde{me3}$$

$$\dot{\tilde{me3}} = m_3 \times \tilde{me2} - d_3 \times \tilde{me3}$$

$$\dot{N} = \frac{\ln(2)}{a + b \times t} \times N,$$

with N the total number of histone tails. Accordingly, the relative H4K20me proportions me0, me1, me2 and me3 are given by $meX = \frac{\tilde{meX}}{N}$, for $X \in \{0, 1, 2, 3\}$ (Alabert et al., 2020) and the corresponding ODEs are given by the chain rule:

$$\dot{meX} = \frac{\dot{\tilde{meX}}}{N} - \frac{\tilde{meX} \times \dot{N}}{N^2}.$$

When plugging in the equations for the absolute H4K20 methylation states into the above equation, we receive the following ODE system for the proportional H4K20 methylation states:

$$\dot{me0} = \frac{-m_1 \times \tilde{me0} + d_1 \times \tilde{me1} + \frac{\ln(2)}{a + b \times t} \times (\tilde{me0} + \tilde{me1} + \tilde{me2} + \tilde{me3})}{N} - \frac{\tilde{me0} \times \frac{\ln(2)}{a + b \times t} \times N}{N^2}$$

$$= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a + b \times t} \times (me0 + me1 + me2 + me3) - me0 \times \frac{\ln(2)}{a + b \times t}$$

$$= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a + b \times t} \times (me1 + me2 + me3)$$

$$\dot{me1} = m_1 \times me0 - \left(m_2 + d_1 + \frac{\ln(2)}{a + b \times t}\right) \times me1 + d_2 \times me2$$

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

CellPress
OPEN ACCESS

$$\dot{me2} = m_2 \times me1 - \left(m_3 + d_2 + \frac{\ln(2)}{a + b \times t}\right) \times me2 + d_3 \times me3$$

$$\dot{me3} = m_3 \times me2 - \left(d_3 + \frac{\ln(2)}{a + b \times t}\right) \times me3$$

$$\dot{N} = \frac{\ln(2)}{a + b \times t} \times N.$$

To constrain the system to biologically meaningful cell-cycle durations we included prior knowledge from literature: at 5.5 hpf the cell-cycle in Xenopus has been found to be ~0.5 h (Heasman, 2006). Hence, we assumed a second linearly increasing cell-cycle function

$$c(t) = 0.5 + b \times t.$$

Mock Model - Scaled Hill function

Similarly, we derive the ODE system of the methylation proportions me0, me1, me2 and me3 for the cell-cycle function $c(t) = a + b \times \frac{t}{h+t}$, a scaled Hill function with Hill coefficient 1 and offset:

$$\dot{me0} = \frac{-m_1 \times \tilde{me0} + d_1 \times \tilde{me1} + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times (\tilde{me0} + \tilde{me1} + \tilde{me2} + \tilde{me3})}{N} \quad - \frac{\tilde{me0} \times \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times N}{N^2}$$

$$= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times (me0 + me1 + me2 + me3) - me0 \times \frac{\ln(2)}{a + \frac{b \times t}{h+t}}$$

$$= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times (me1 + me2 + me3)$$

$$\dot{me1} = m_1 \times me0 - \left(m_2 + d_1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}}\right) \times me1 + d_2 \times me2$$

$$\dot{me2} = m_2 \times me1 - \left(m_3 + d_2 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}}\right) \times me2 + d_3 \times me3$$

$$\dot{me3} = m_3 \times me2 - \left(d_3 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}}\right) \times me3$$

$$\dot{N} = \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times N.$$

Similar to the mock model with linearly increasing cell-cycle function we tested three different scaled Hill functions with Hill coefficient 1 and offset as cell-cycle functions:

$$c(t) = a + \frac{b \times t}{h+t}$$

$$c(t) = 0.5 + \frac{b \times t}{h + t}$$

$$c(t) = 0.5 + \frac{b \times t}{b + t}.$$

We again reduced the number of model parameters in the second equation by inserting prior knowledge about the cell-cycle duration at the start of the model (see Mock Model - Linearly Increasing Cell-Cycle Duration). Additionally, we reduced the number of model parameters further by assuming the scaling b and the dissociation constant h to be identical in the third equation. In comparison to the former two cell-cycle functions, the third equation constrains the parameter space more strictly. E.g. for upper and lower boundaries of 100 and 0.0001 for parameters a, b and h the first equation will allow for cell-cycle durations up to 100+100 × (42-5.5)/(0.0001+42-5.5) ≈ 200 hours while the third equation only allows for cell-cycle durations for up to 0.5+100 × (42-5.5)/(100+42-5.5) ≈ 27 hours. Additionally, we tested whether inferring the Hill coefficient with the other parameters rather than fixing it to 1 would lead to similar or improved results. Hill functions with Hill coefficient > 1 first increase slowly before a rapid increase and a gradual plateau follows. This does not reflect the biologically observed cell-cycle dynamics. Hill coefficients $\leq$ 1 lead to a fast initial increase which could reflect known cell-cycle dynamics. We ran the optimization for the best performing mock model without demethylation this time inferring the Hill coefficient with the other parameters (lower and upper boundaries of 0.001 and 1, respectively). We found the inferred Hill coefficient to be 0.5. While the BIC values for both models are comparable (-23 and -21 for the models with Hill coefficient = 1 and inferred Hill coefficient, respectively) the average cell-cycle durations differ (8 h and 15 h, respectively) (Table S1). By choosing a Hill coefficient = 1, we receive biologically meaningful average cell-cycle durations while reducing the number of inferred parameters by 1 and maintaining the same goodness of fit. Hence, all analyses were performed using the scaled Hill function with Hill coefficient 1 and offset 0.5.

Noise Models

As experimental data is generally noise corrupted, we evaluated all models with an underlying Laplacian noise model. Maier et. al. (Maier et al., 2017) have shown that Laplacian noise models may outperform Gaussian ones due to their increased robustness against outliers (Maier et al., 2017). All model parameters are comprised in the parameter vector θ and the experimental measurement i at time point k is denoted by $\bar{y}_i^k$. The log-likelihood for the Laplacian noise model is given by

$$\log L(\theta) = -\sum_{i,k}\left(\log(2\sigma) + \frac{\log(\bar{y}_i^k) - \log(y_i(t_k,\theta))}{\sigma}\right).$$

By performing maximum likelihood estimation we obtain the optimal model parameters.

OPTIMIZATION AND PARAMETER ESTIMATION

The model parameters include the initial proportions, $me1_0$, $me2_0$ and $me3_0$. Without loss of generality, we fix relative initial proportion $me0_{0rel}$=0.1 to obtain structural identifiability, where the relative initial proportions are given by

$$meX_0 = \frac{meX_{0rel}}{\sum_{i=0}^{3} mei_0},$$

with X ∈ {0, 1, 2, 3} and meX_{0rel} the relative initial proportions. We additionally infer one noise parameter, the model-specific rate constants of (de-)methylation and potentially up to three constants (mock models) describing the cell-cycle function. We tested whether the fixation of the relative initial proportion $me0_{0rel}$ influences the robustness of the optimization by fixing $me0_{0rel}$=0.01 and $me0_{0rel}$=1 for the best performing mock model without demethylation. We found that the optimized rate constant parameter sets are robust to the initializations of the relative initial proportion of unmethylated H4K20 (see Table S1). For numerical reasons we optimized the parameters in a $\log_{10}$ scale (Hass et al., 2019). The lower and upper bounds for the rate constants, initial states, noise parameter and cell-cycle parameters were initiated in $\log_{10}$ scale at -10 to 2, -4 to 2, -2 to 0 and -10 to 10, respectively. We performed multi-start local optimization of the negative log-likelihood using the parameter estimation toolbox PESTO (Stapor et al., 2018) and simulated the models with AMICI (Fröhlich et al., 2017). We performed at least 100 local optimization runs per model, initialized by latin-hyper cube-sampled starts. For the models not converging upon these initializations (where by 'not converging' we mean that the likelihood value of the second best run differs more than 0.1) we decreased the width between upper and lower bounds to increase the probability of convergence. For this, we assured that the optimization bounds were wide enough such that the optimal values are not in the bounds for the rate constants and the cell-cycle parameters. As the initial states are unidentifiable we ignored optimal values which ran into these boundaries as long as other optimal values were found within. For models where this was not the case we expanded the boundaries of the rate constants and initial states up to -20 to 10 and -10 to 10,

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

respectively, as we assumed any smaller or larger values to be biologically non-informative. For the initially non-converged joint models we also increased the number of starts to 800. Models not having converged upon manually adjusting the boundaries and running for 800 starts were determined to not converge. All mock and HUA models converged. We determined 5 out of the 40 joint models to not converge (Table S1). The given likelihood values of these joint models are lower bounds of the true optimal likelihood values obtainable upon convergence. As the likelihood values of all 5 non-converged joint models still resulted in considerably lower BIC values in comparison to the other tested models we can safely report them as best performing models for the respective demethylation hypothesis. As the BIC values between the demethylation hypotheses allowing and not allowing for demethylation in HUA differ considerably we assume the comparison between different demethylation hypotheses to be valid and the resulting conclusions to be justified.

Model Selection

We use the Bayesian Information Criterion (BIC) (Schwarz, 1978) for model comparisons:

$$\mathrm{BIC} = \ln(n) \times k - 2 \times \log L,$$

where n is the number of data points, k is the number of estimated parameters or the overall model complexity and $\log L$ is the log-likelihood value for the maximum likelihood estimate of the model parameters. The BIC rewards high likelihood values and penalizes model complexity. Hence, low BIC values are preferable. In comparison to other model selection methods such as the Akaike Information Criterion (AIC) the BIC penalizes additional model complexity more strongly. We consider a $\Delta BIC > 10$ between two models to be enough evidence to reject the model with the higher BIC (Kass and Raftery, 1995).

Parameter Uncertainty

To receive the uncertainties for the estimated model parameters we performed Markov Chain Monte Carlo (MCMC) sampling of the posterior distribution

$$p(\theta|D) \propto L(\theta)\pi(\theta),$$

with uniform prior $\pi(\theta)$ defined over the optimization boundaries, likelihood function $L(\theta)$ and data D. We sampled the posterior for all six best performing joint models and the mock model with three specific methylation rate constants and no demethylation (PESTO-internal function *getParameterSamples*). We employed parallel tempering with five parallel chains initiated at the five most optimal parameter estimates per model obtained during optimization and performed 10^6 iterations. Upon performing a Geweke test (first 10% versus last 50% of the final MCMC chains), we discarded the first 10% of the samples as burn-in phase and thinned the chains keeping only every 100th sample. The marginal posterior distributions are plotted via violin plots (plotting function *violin*, Hoffman, H. (2015). violin.m - Simple violin plot using matlab default kernel density estimation. (https://de.mathworks.com/matlabcentral/fileexchange/45134-violin-plot), MATLAB Central File Exchange. Retrieved November 13, 2019.)).

Validation - Cell-Cycle Durations

We used the median and the 25th and 75th percentiles of the MCMC chain determined during the parameter uncertainty analysis for the cell-cycle parameter b, and evaluated the median and the 25th and 75th percentiles of the cell-cycle function according to

$$c(t) = 0.5 + \frac{b \times t}{b + t},$$

for $t \in [0,40]$, where the cell-cycle duration of 0.5 hours at 5.5 hpf (start of model) is taken from (Anderson et al., 2017; Gelens et al., 2015).

Prediction of Number of Cells

Using the median and the 25th and 75th percentiles of the cell-cycle parameter b (as determined in the validation analysis), we determined the theoretical number of cells a Xenopus embryo is on average composed of between 5.5 hpf and 45.5 hpf according to dN/dt = ln(2)/(0.5+b*t/(b+t))*N, where

$$\frac{\partial N(t)}{\partial t} = \frac{\ln(2)}{0.5 + \frac{b \times t}{b + t}} \times N$$

$$N(t) = N0 \times e^{(2 \times \ln(2))/(2 \times b + 1)^2 \times \left(2 \times b^2 \times \ln(2 \times b \times t + b + t) + 2 \times b \times t + t - 2 \times b^2 \times \ln(b)\right)},$$

where N(t) is the number of cells at time t, N0 the initial number of cells and 0.5+b*t/(b+t) the cell-cycle function (constrained scaled Hill function with Hill coefficient 1 and offset 0.5). At the start of the model (at 5.5 hpf) we take the initial number of cells N0 to be 4096 (Heasman, 2006).

Please cite this article in press as: Schuh et al., H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.11.003

Implementation

The toolboxes used for the analysis of the manuscript for ODE simulation (AMICI (Fröhlich et al., 2017)) and parameter estimation (PESTO (Stapor et al., 2018)) are available under https://github.com/ICB-DCM. The MATLAB code corresponding to this manuscript is available via https://github.com/marrlab/HistonesXenopus. The analysis was performed with MATLAB 2017a.

QUANTIFICATION AND STATISTICAL ANALYSIS

For comparing H4K20me data for mock and HUA (Figure 1B), a two-sample t-test at the 0.05 significance level was used for all three biological replicates of mock and HUA for each time point. In Figures 2F and 2G, the model predictions are given for the median, 25^{th} and 75^{th} percentiles of the MCMC samples of the cell-cycle parameter of the model with three specific methylation rate constants but no demethylation. Stated values of specific (de-)methylation rate constants are given by the median and the credibility ranges from the 25^{th} to the 75^{th} percentiles. All statistics and analyses were performed with MATLAB 2017a.

B. Further manuscript as main author currently under review

B.1. Gal1 repression memory in budding yeast.

Gal1 repression memory in budding yeast.

Lea Schuh, Igor Kukhtevich, Poonam Bheda, Melanie Schulz, Maria Bordukova, Robert Schneider, and Carsten Marr.

Transcriptional adaptation to changing environments is crucial for cell survival. When exposed e.g., to repeated stimulation cells can 'remember' their initial transcriptional response leading to an adapted reaction upon re-stimulation, a phenomenon termed transcriptional memory. Most studies on transcriptional memory so far have focused on gene induction, although gene repression plays a similarly central role in gene regulation. This opens the important question as to whether memory also exists for repression and how repession memory would manifest at a single-cell level. To identify whether there is repression memory in budding yeast cells, we compared the repression kinetics of Galactokinase 1 (Gal1) expression of two consecutive repression periods. First, we developed a method to compensate for dilution effects due to budding, the developing of a new cell. This enabled me to deconvolute dilution and repression kinetics. Next, we formulated two mathematical models based on ordinary differential equations to appropriately describe the repression kinetics of repressor and non-repressor cells. While the non-repressor cells are assumed to demonstrate constant basal Gal1 production, the repressor cells are assumed to first actively produce Gal1 until they re-establish delayed repression and switch off Gal1 production. For both models, we determined the analytical solutions. We performed multi-start maximum likelihood optimization and model selection using the Bayesian Information Criterion on each of the single-cell Gal1 repression traces for each repression period. This enabled me to quantitatively describe the single-cell repression kinetics and to discriminate between repressor and non-repressor cells. Using the estimated parameters of the repressor cells, we identified that the repression delay, i.e., the time between carbon source switch and termination of Gal1 expression, is statistically significantly decreased in cells during repeated repression, suggesting the existence of repression memory. Finally, we applied the same modeling approach to the Gal1 repression kinetics of a mutant, *elp6*Δ and identified *elp6*Δ as a novel gain-of-repression-memory mutant exhibiting a stronger decrease in the repression delay between first and second repression.

Statement of individual contribution

This work was motivated by several discussions between Robert Schneider, Carsten Marr and myself. I had the idea to model the Gal1 repression kinetics of every single-cell and repression period individually in order to receive single-cell parameters and to quantitatively describe and compare single-cell repression kinetics. I subsequently formulated the mathematical models for the repressor and non-repressor cells, developed a method to compensate for dilution effects in the Gal1 kinetics and performed parameter estimation and model selection on the single-cell Gal1 traces. Moreover, I performed the statistical analysis on the estimated parameters of the repressor cells and applied the same modeling approach to the Gal1 repression kinetics of the *elp6*Δ mutant. I was responsible for the entire methodology, formal analysis and writing of the manuscript.

I, Lea Schuh, am the main author of this manuscript.

Gal1 repression memory in budding yeast

Lea Schuh[1,2,3,4], Igor Kukhtevich[2,] Poonam Bheda[2], Melanie Schulz[1,3], Maria Bordukova[3], Robert Schneider[2,5,6,*], and Carsten Marr[1,3,*]

[1]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany
[2]Institute of Functional Epigenetics, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany
[3]Institute of AI for Health, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany
[4]Department of Mathematics, Technical University of Munich, Garching, 85748, Germany
[5]German Center for Diabetes Research (DZD), 85764 Neuherberg, Germany
[6]Faculty of Biology, Ludwig-Maximilians University of Munich, 82152 Planegg-Martinsried, Germany
*correspondence: robert.schneider@helmholtz-muenchen.de, carsten.marr@helmholtz-muenchen.de

ABSTRACT

Cells must continuously adapt to changing environments and, thus, have evolved mechanisms allowing them to respond to repeated stimuli. For example, faster gene induction upon a repeated stimulus aids adaptation - a process known as reinduction memory. However, whether such a memory exists for gene repression is unclear. Here, we studied gene repression across repeated carbon source shifts in over 2,500 single *Saccharomyces cerevisiae* cells. By monitoring the expression of a carbon source-responsive gene, galactokinase 1 (*Gal1*), and mathematical modeling, we discovered repression memory at the population and single-cell level. Using a repressor model to estimate single-cell repression parameters, we show that repression memory is due to a shortened repression delay, the estimated time gap between carbon source shift and Gal1 expression termination, upon the repeated carbon source shift. Additionally, we show that cells lacking *Elp6* display a gain-of-repression-memory phenotype characterized by a stronger decrease in repression delay between two consecutive carbon source shifts. Collectively, our study provides the first quantitative description of repression memory in single cells.

INTRODUCTION

Cells receive and process external signals to optimally adapt to changing environments. Repeated stimulation from the same external signal induces an adapted transcriptional response, a phenomenon termed transcriptional memory (*1*). It is crucial to understand the mechanisms underlying transcriptional memory due to its implications for a broad range of cellular functions, including the human adaptive immune system (*2*, *3*), disease development in diabetes (*4*, *5*), and aging (*6*). However, transcriptional memory has primarily been researched concerning gene induction despite gene repression playing an essential role in gene regulation (*7*, *8*). This raises the question of whether memory exists in repression.

The adaptation of *Saccharomyces cerevisiae* (budding yeast) to carbon sources is among the most well-studied eukaryotic signal integration systems. Whereas glucose directly enters glycolysis, a vital metabolic route providing cells with energy, galactose is first converted to glucose-6-phosphate (*9*, *10*), necessitating the production of Gal gene-encoded enzymes (*11*). Repeated alternations between glucose and galactose media revealed that yeast cells are primed by their carbon source history, exhibiting transcriptional memory: repeated galactose induction results in enhanced Gal gene expression (*12*–*16*). Bheda et al. examined the expression of galactokinase 1 (*Gal1*) in single cells, for which reinduction memory has been well characterized, and discovered that a shorter delay, rather than an increased expression rate, contributed to the observed increase in Gal1 levels (*17*). Moreover, they identified *elp6Δ* as a gain-of-reinduction memory mutant, with *elp6Δ* cells showing Gal1 levels comparable to wildtype cells in the first induction, but earlier induction onset and increased Gal1 levels in the second induction. While, as these examples show, reinduction memory upon galactose induction has been thoroughly researched, and, while, Lee et al. suggested repression memory for bulk populations (*18*), it is uncertain if individual cells display Gal1 repression memory upon repeated glucose repression (Figure 1).

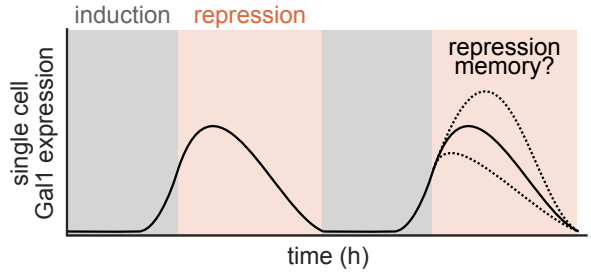


Figure 1. Does gene repression memory exist in budding yeast?
Previous studies focused on gene expression during induction. However, whether there is memory in repression indicated by altered repression kinetics is still unknown.

We measured Gal1 expression in wildtype budding yeast cells via a Gal1-GFP (green fluorescent protein) fusion across repeated galactose inductions and glucose repressions using time-lapse microscopy coupled with a microfluidic device to follow and study Gal1 repression kinetics. In the second repression, we discovered that the time of maximal mean Gal1 expression was shortened. By using a mathematical model to quantify single-cell Gal1-GFP kinetics during glucose repression, and by distinguishing between repressor and non-repressor cells, we revealed that the shortened time to maximal mean Gal1 expression at the population level was not caused by different fractions of repressor cells between consecutive repressions. Using the estimated single-cell parameters, we found the repression delay, which is the estimated time gap between the galactose to glucose shift and the Gal1 expression termination, to be shortened in the second repression at the single-cell level, implying Gal1 repression memory. Furthermore, we repeated the experiments and analysis for the gain-of-induction memory mutant *elp6Δ.* Remarkably, *elp6Δ* cells showed a stronger repression memory effect than wildtype, making *elp6Δ* also a gain-of-repression memory mutant.

RESULTS

Automated time-lapse microscopy and microfluidics allow for the quantification of single-cell Gal1 repression kinetics across repeated carbon source shifts

To study Gal1 repression kinetics over multiple repressions, we exposed wildtype budding yeast cells alternatingly to glucose or galactose media (Figure 2A). For this, we cultured the cells in custom-made microfluidic devices to ensure precise media shifts and long-term tracking (*17*). Gal1 expression levels in single yeast cells were monitored using a Gal1-GFP fusion, a standard reporter to study gene expression in time-lapse microscopy (see Bheda et al., 2020 for details). We captured images from the microfluidics chambers every 3 min totaling 320 images per chamber during a 16-h experiment. The yeast cells were then semi-automatically segmented, mapped and the total Gal1-GFP fluorescence signal per cell and time point was extracted using Autotrack and PhyloCell (*19*), YeaZ (*20*) and Cell-ACDC (*21*) (see Materials and Methods), yielding over 2,500 single-cell Gal1 expression traces (Figure 2B). Asymmetric budding allowed us to identify mother-daughter relationships. As expected, the total Gal1-GFP fluorescence signal of single cells reveals Gal1 inductions and repressions during galactose and glucose, respectively, and increased overall Gal1 levels in induction i2 (Figure 2B).

During budding, cytoplasmic proteins are disseminated between the mother and daughter cells. Assuming a constant Gal1 protein amount, its redistribution decreases the total Gal1-GFP fluorescence signal in the mother cell (Figure 2C top), a phenomenon called dilution. To deconvolute dilution and repression kinetics, we calculated the sum of the total Gal1-GFP fluorescence signal of the mother cell and its progeny (Figure 2C bottom, and see Materials and Methods). In the following, the adjusted sum of the total Gal1-GFP fluorescence signal of the mother cell and its progeny is referred to as total GFP. We applied the same dilution compensation to repressions r1 and r2 (Figure 2D).

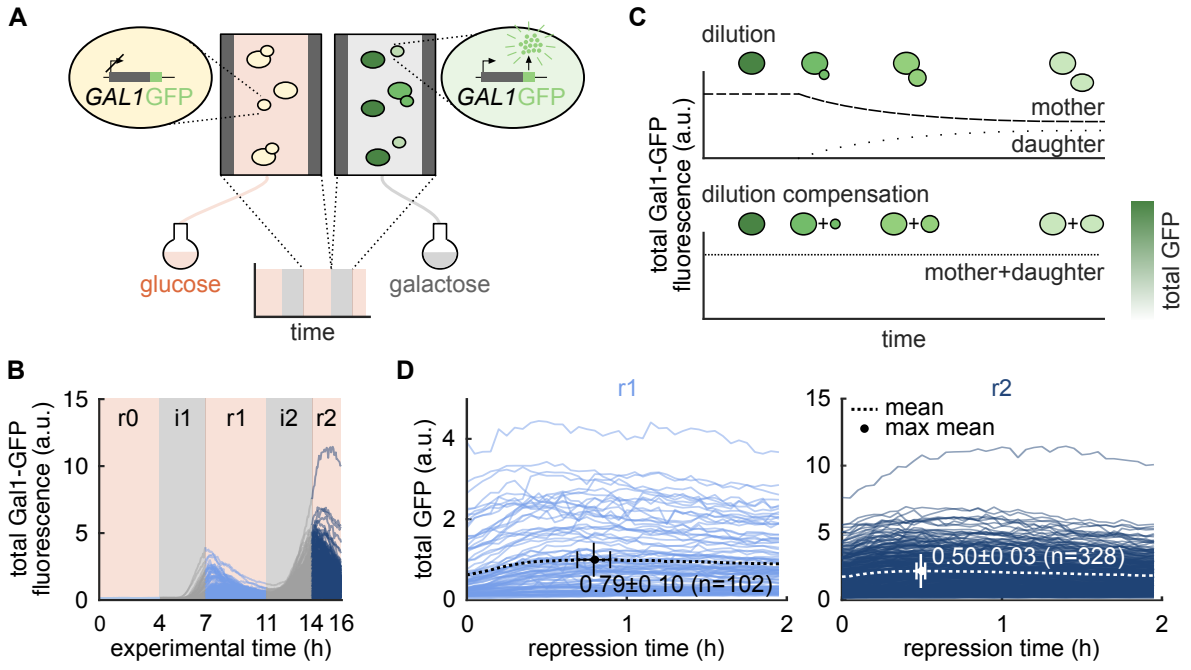


Figure 2. A shortened time to maximal mean total GFP in the second repression at the population level.
(A) Budding yeast cells were grown in microfluidic chambers and alternatingly exposed to a medium containing either glucose (orange) or galactose (gray) as carbon source. Galactokinase 1 (*Gal1*) is induced in cells exposed to galactose and repressed in cells exposed to glucose. Gal1 expression was monitored via a Gal1-GFP fusion and time-lapse microscopy.
(B) Single-cell traces of total Gal1-GFP fluorescence signal across two inductions i1 and i2 (gray) and repressions r0, r1, and r2 (blue).
(C) Budding decreases the total Gal1-GFP fluorescence signal in mother cells and increases the total Gal1-GFP fluorescence signal in daughter cells (top). To compensate for this dilution, we summed up the total Gal1-GFP fluorescence signal of each mother cell present and its progeny during one repression (bottom).
(D) Single-cell traces of total GFP signal adjusted for dilution (see (C)) for the first two hours of repressions r1 (left) and r2 (right). Time to maximal mean total GFP is 17 min shorter in repression r2, where mean total GFP is indicated by the dotted line and the maximal mean total GFP is highlighted by the dot. Bootstrap (10^5) samples were drawn to generate mean ± std.

Decreased time to maximal total GFP in repression r2 at the population level

Following a galactose-glucose shift, total GFP intensities initially rise before decreasing (Figure 2D). To determine the repression kinetics at the population level, we calculated the mean total GFP signal over time for repressions r1 and r2. Interestingly, the time to attain the maximal mean total GFP reduced from 0.79 ± 0.10 h (mean ± std, n = 102 cells) in r1 to 0.50 ± 0.03 h (n = 328 cells) in r2, where the time point of the maximal mean was bootstrapped 10^5 times (Figure 2D). This demonstrates a decreased time to maximal total GFP in r2 at the population level.

The computational model distinguishes between repressor and non-repressor cells

At the single-cell level, Gal1 induction delay varies significantly. Zacharioudakis et al. showed that Gal1 induction caused by a glucose–galactose media shift results in a bimodal population distribution, with only a subset of cells inducing Gal1 even after several hours of galactose exposure (*16*). As in our experiments, repression was preceded by 3 h of galactose induction, we expected that our cell population at the start of repression contained induced and uninduced cells, which show and do not show repression kinetics, respectively. However, independent of repression memory, a larger proportion of non-repressor cells in repression r1 could explain the decrease in the time to maximal mean total GFP in repression r2 (Figure 2D). As it is difficult to distinguish between repressor and non-repressor cells from the total GFP traces alone (Figure 2D), we used computational modeling and model selection to systematically describe the kinetics of single total GFP traces. Since Gal1 induction results in an approximate 1000-fold change in Gal1 expression (*22*), we assumed that stochasticity inherent to gene expression was insignificant and that a deterministic modeling approach was sufficient in describing the kinetics of the total GFP traces. To discriminate between repressor and non-repressor cells, we defined two models. The non-repressor model assumes a constant basal GFP production and degradation over time with rates r_{basal} and r_{deg} (Figure 3A left), since we observed a gradual increase in total GFP signal in cells visually identified as not showing repression kinetics (Figure 3B top right). The temporal variation of the total GFP signal over time indicated by the non-repressor model is summarized using the following ordinary differential equation:

$$dGFP(t)/dt = r_{basal} - r_{deg}\, GFP(t) .$$

This is solved by

$$GFP(t) = r_{basal}/r_{deg}\,(1 - \exp(-r_{deg}t)) + GFP_0\exp(-r_{deg}t),$$

where $GFP_0 = GFP(0)$, the initial total GFP at time point 0. According to the repressor model, cells that induced Gal1 during galactose induction required time to reestablish glucose-mediated repression. Hence, GFP is actively generated at rate r_{prod} till a time point t_{delay}. GFP production is switched off ($r_{prod} = 0$) and GFP is degraded with rate r_{deg} after this estimated repression delay t_{delay} (Figure 3A right). An example of a cell visually identified as showing repression kinetics can be found in Figure 3B top left. Until t_{delay}, the repressor model equals the non-repressor model. The temporal change of total GFP over time described by the repressor model is summarized by the following ordinary differential equations:

$$\text{before } t_{delay}\text{: } dGFP(t)/dt = r_{prod} - r_{deg}\, GFP(t)$$

$$\text{after } t_{delay}\text{: } dGFP(t)/dt = -r_{deg}\, GFP(t)$$

with solutions

$$\text{before } t_{delay}\text{: } GFP(t) = r_{prod}/r_{deg}\,(1 - \exp(-r_{deg}t)) + GFP_0\exp(-r_{deg}t)$$

$$\text{after } t_{delay}\text{: } GFP(t) = GFP(t_{delay})\exp(-r_{deg}(t - t_{delay})),$$

where $GFP(t_{delay}) = r_{prod}/r_{deg}\,(1 - \exp(-r_{deg}t_{delay})) + GFP_0\exp(-r_{deg}t_{delay})$. The non-repressor and repressor model comprise four and five model parameters, respectively: initial total GFP, GFP_0, basal GFP production rate, r_{basal}, or GFP production rate, r_{prod}, GFP degradation rate r_{deg}, a noise parameter σ determining the width of the Gaussian noise distribution (see Materials and

Methods), and the repression delay t_{delay} for the repressor model. For repressions r1 and r2, respectively, we performed multi-start maximum likelihood optimization and model selection on both models for each total GFP trace (Figure 3B top and 3C). Calculating the profile likelihoods of exemplary total GFP traces, we found the model parameters of the repressor model to be identifiable (Figure 3B bottom). We then determined whether active repression, i.e. the repressor model, was required to explain a total GFP trace using the Bayesian information criterion (BIC). A BIC difference of ten between the repressor and non-repressor model ($BIC_{repressor} < BIC_{non\text{-}repressor} - 10$) was considered an appropriate threshold to reject the non-repressor model with fewer model parameters (see Materials and Methods and Figure 3B). Of all total GFP traces, 71% and 76% of r1 and r2, respectively, were discovered to require the repressor model (Figure 3D) and are henceforth referred to as "repressor cells." The higher proportion of repressor cells was expected due to Gal1 transcriptional reinduction memory resulting in increased proportions of cells producing detectable GFP in induction i2. The median initial total GFP, GFP_0, was significantly higher in repressor cells than in non-repressor cells ($p = 9.2 \cdot 10^{-5}$ for r1 and $p = 2.7 \cdot 10^{-7}$ for r2) (Figure 3D). This implies that our repression models can discriminate between cells that were repressing Gal1 and cells uninduced at the beginning of repression. The overlapping ranges of initial total GFP between repressor and non-repressor cells reveal how simple thresholding could result in wrong differentiation between repressor and non-repressor cells.

Decreased time to maximal total GFP is also present in the repressor cell subpopulation

To determine whether the previously described decrease in time to maximal mean total GFP in r2 is due to a different fraction of non-repressor cells between r1 and r2, we computed the times to maximal mean total GFP on the repressor cell subpopulation. We again found a shortened time to maximal mean total GFP in r2, with 0.81 ± 0.09 (mean ± std, n = 72 cells) and 0.52 ± 0.05 (n = 248 cells) h for r1 and r2, respectively (Figure 3E), demonstrating that the decrease in time to maximal total GFP in r2 is not due to a different fraction of non-repressor cells.

Shortened repression delay in repression r2 at the single-cell level

To address if the repression kinetics are different between r1 and r2 in individual cells, we compared the paired estimated single-cell parameters of repressor cells present in both repressions. We discovered that the median initial total GFP, GFP_0, and median repression delay, t_{delay}, are substantially different ($p = 3.3 \cdot 10^{-9}$ and $p = 1.5 \cdot 10^{-3}$, respectively) between both repressions using a two-sided paired sign test and multiple test correction (Figure 3F). Median GFP_0 is increased (median values of 0.59 and 2.05 for r1 and r2, respectively), while median t_{delay} is shortened in r2 (median values of 0.50 and 0.38 for r1 and r2, respectively, where 72% of paired cells showed a decrease in t_{delay}), in line with the previously identified decrease in the time to the maximal mean total GFP in r2 at the repressor subpopulation level (Figure 3E). The increased GFP_0 in r2 conforms to the transcriptional reinduction memory of Gal1 and results in higher GFP_0 at the start of r2. The median r_{prod} and median r_{deg} between the two repressions were comparable

(p = 0.22 and p = 0.34, respectively) (Figure 3F). In line with our findings, Bheda et al. also identified similar production rates (*17*). We repeated the entire wildtype analysis based on data from independent experiments (see Methods for details). This replicate analysis confirms the conclusions presented in Figures 2B, 2D and 3 , in particular the earlier repression response in repression r2 (Figure S1).

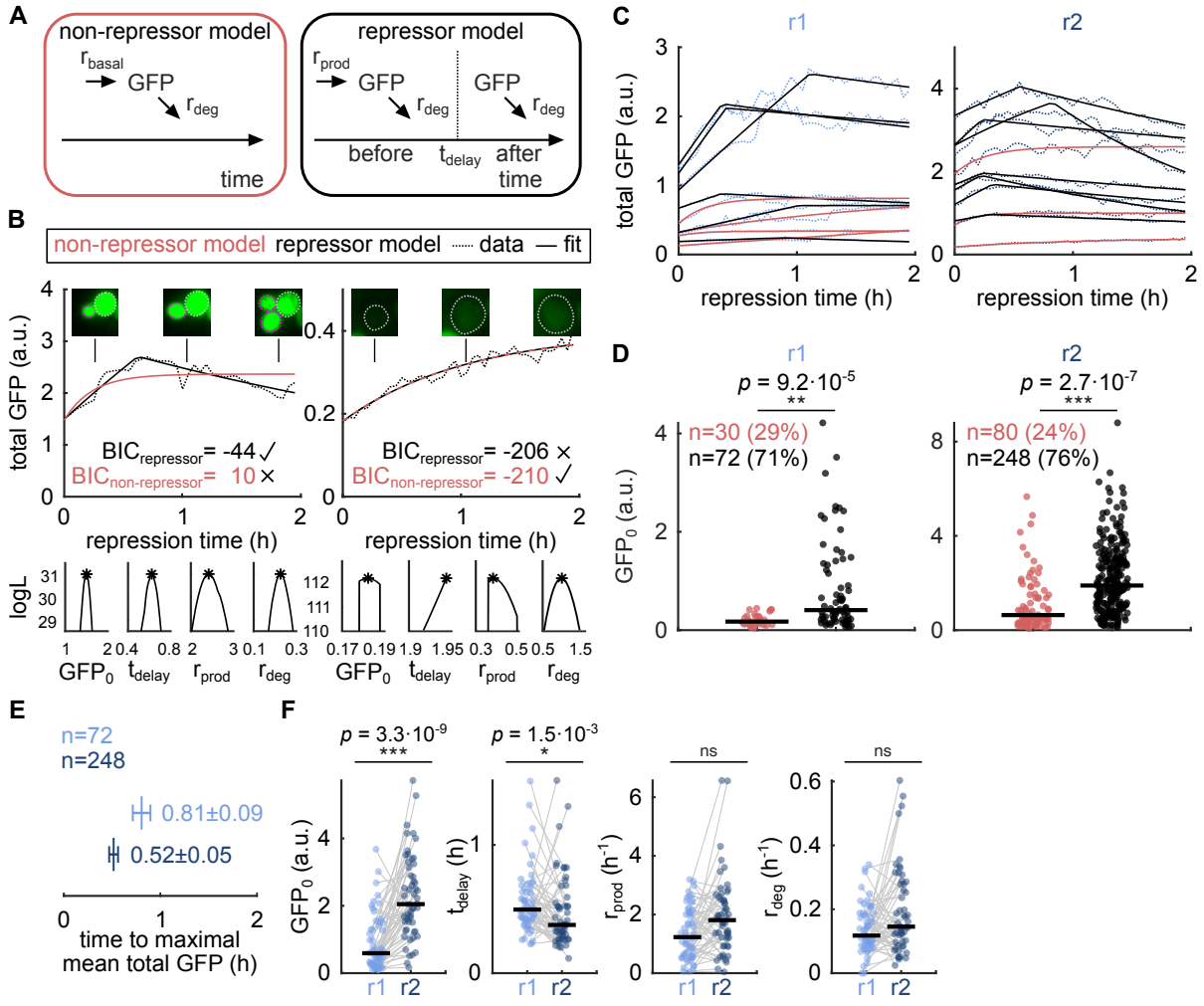


Figure 3. Shortened repression delay in repression r2 at the single-cell level.
(A) Left: a model for non-repressor cells composed of basal GFP production (r_{basal}) and degradation (r_{deg}). Right: a model for repressor cells composed of an initial constant and active GFP production (r_{prod}) and degradation (r_{deg}) until a delayed repression onset (t_{delay}) where GFP production is switched off.
(B) Top: two exemplary total GFP traces (dotted line) and fits of the non-repressor model (red solid line) and repressor model (black solid line). Exemplary images of the cell(s) at three different time points are shown above. Mother cells are circled in gray, progeny in pink. The better fitting model was selected according to the Bayesian information criterion (BIC). Left: total GFP trace better fitted by the repressor model. Right: total GFP trace fitted equally well by the non-repressor and repressor model. Due to the higher model complexity of the repressor model, the repressor model is still rejected. Bottom: profile

likelihoods of the repressor model corresponding to the two exemplary total GFP traces above endorse parameter identifiability. Asterisks represent optimized parameters and corresponding log-likelihood (logL) values.
(C) Ten exemplary total GFP traces (dotted lines) and best fits (solid lines) for repressions r1 (left) and r2 (right). GFP traces best fitted with a repressor model are shown in black and fits of total GFP traces best fitted with a non-repressor model are shown in red.
(D) The median initial total GFP, GFP_0, is significantly higher ($p = 9.2 \cdot 10^{-5}$ and $p = 2.7 \cdot 10^{-7}$, two-sided median test corrected for multiple testing with Bonferroni correction, m = 20) in traces better fitted by the repressor model (black) than in traces better fitted by the non-repressor model (red). This confirms that the repressor model fits induced cells better, while the non-repressor model fits uninduced cells. The number of cells and percentages of all GFP traces best fitted by the repressor model and non-repressor model are shown.
(E) Time to maximal mean total GFP is decreased in repression r2 for repressor cells (0.81 ± 0.09 vs. 0.52 ± 0.05). Bootstrap (10^5) samples of the repressor cells were drawn to generate mean ± std.
(F) Comparison of paired estimated single-cell parameters of repression r1 and r2 shows that the median initial total GFP, GFP_0, and median repression delay, t_{delay}, are significantly different ($p = 3.3 \cdot 10^{-9}$ and $p = 1.5 \cdot 10^{-3}$, respectively, two-sided paired sign test correcting for multiple testing with Bonferroni correction, m = 20, and the number of paired cells = 54), with median GFP_0 increased and median t_{delay} decreased (median t_{delay} values of 0.50 and 0.38 for r1 and r2, respectively).

Earlier repression response in repression r2 for *elp6Δ* cells

Intrigued by the findings in wildtype yeast cells, we repeated our analysis for the previously identified gain-of-reinduction memory mutant *Elp6* (*elp6Δ*). The total Gal1-GFP fluorescence signal of single *elp6Δ* cells demonstrates Gal1 induction and repression during galactose and glucose and reinduction memory (Figure 4A). We performed dilution compensation (Figure 2C) on repressions r1 and r2 (Figure 4B) and calculated the mean total GFP signals over time to determine the repression kinetics at the population level. Similar to wildtype cells, we discovered that the time to attain the maximal mean total GFP decreased from 1.40 ± 0.09 h (n = 66 cells) in r1 to 0.71 ± 0.07 h (n = 237 cells) in r2, (Figure 4B). Subsequently, we repeated multi-start maximum likelihood optimization and model selection for the non-repressor and repressor models to show that the shortened time to maximal mean total GFP in r2 is not caused by a larger proportion of non-repressor cells in r1 (Figure 4C). Of all total GFP traces, 64% and 83% of *elp6Δ* cells in r1 and r2 require the repressor model (Figure 4D). When determining the time to maximal mean total GFP for the *elp6Δ* repressor subpopulation, we again identified a shortened time to maximal mean total GFP, with 1.20 ± 0.14 (n = 42 cells) and 0.69 ± 0.07 (n = 196 cells) h for r1 and r2, respectively (Figure 4E). To investigate the earlier repression response at the single-cell level, we compared the paired estimated single-cell parameters of *elp6Δ* repressor cells present in both repressions. We found the median GFP_0, and median t_{delay}, to be significantly different ($p = 1.2 \cdot 10^{-10}$ and $p = 3.9 \cdot 10^{-5}$) between repressions (Figure 4F). Similar to wildtype cells, the median GFP_0 increased (median values of 0.57 and 3.69 for r1 and r2, respectively), while the median t_{delay} is shortened in r2 (median values of 0.95 and 0.50 for r1 and r2, respectively, where 85% of paired cells showed a decrease in t_{delay}), as previously identified at the repressor subpopulation level. The increased initial total GFP level in r2 was expected due to the previously identified reinduction memory of *elp6Δ* cells. Contrary to wildtype cells, we discovered that for *elp6Δ* cells the median production rate, r_{prod}, is significantly different between r1 and r2 ($p = 6.9 \cdot 10^{-8}$) with increased median r_{prod} in r2 (median values of 1.10 and 3.10 for r1 and r2,

respectively, Figure 4F) reflecting the gain-of-reinduction-memory phenotype reported by Bheda et al. (*17*). Median r_{deg} was comparable between repressions (p = 0.39). We repeated the *elp6Δ* analysis based on additional data from independent experiments (see Methods for details). The replicate analysis confirms the conclusions presented in Figure 4, in particular the earlier repression response in repression r2 (Figure S2).

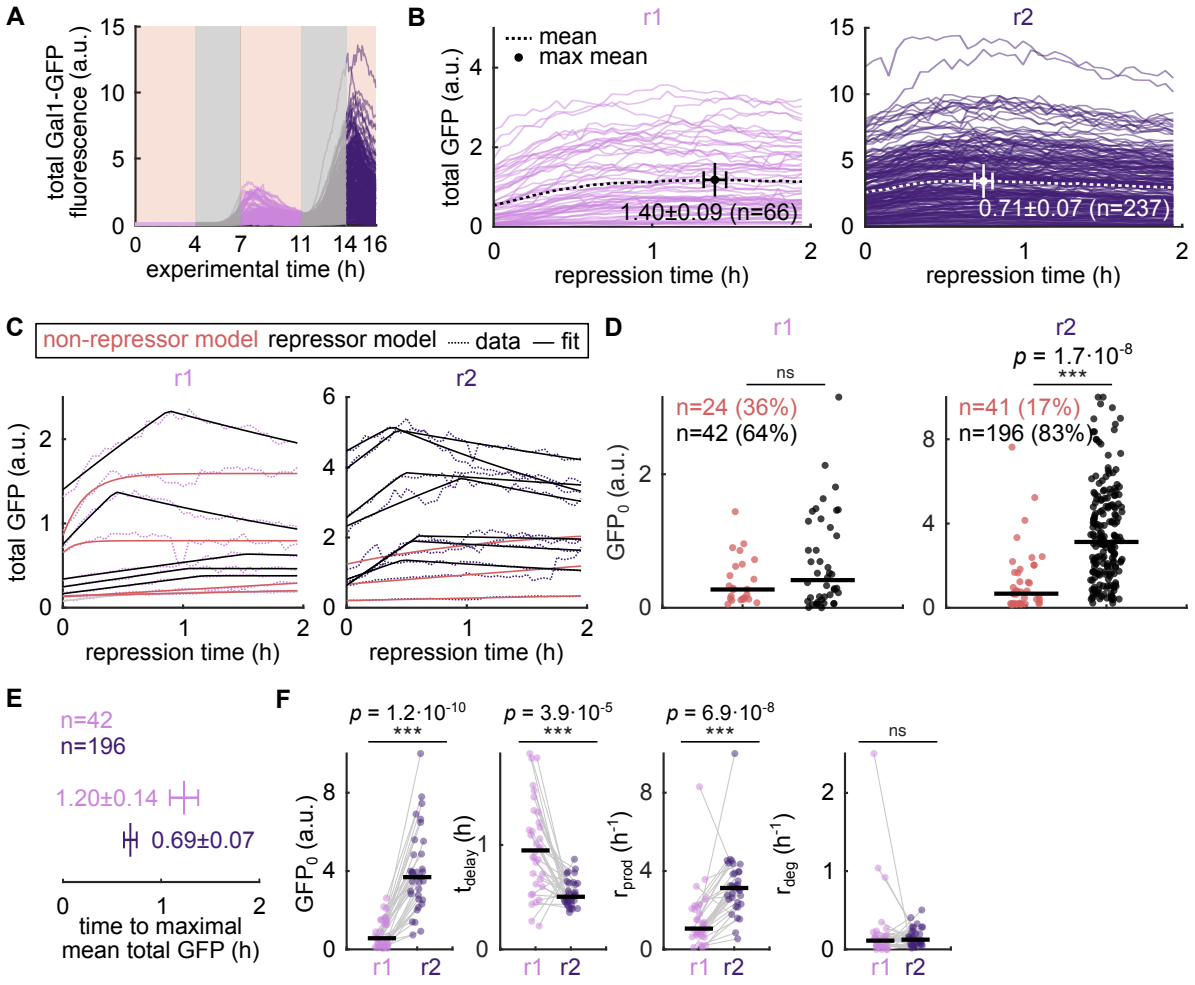


Figure 4. Earlier repression response in repression r2 for *elp6Δ* cells at both the population and single-cell level.
(A) Single-cell traces of total Gal1-GFP fluorescence signal of *elp6Δ* budding yeast cells across two inductions i1 and i2 (gray) and repressions 0, 1, and 2 (purple).
(B) Single-cell traces of total GFP signal of *elp6Δ* budding yeast cells adjusted for dilution (see Figure 2C) for the first two hours of repression 1 (r1, left) and repression 2 (r2, right). Time to maximal mean total GFP is 41 min shorter in repression r2, where the mean total GFP signal is indicated by the dotted line and the maximal mean total GFP is highlighted by the dot. Bootstrap (10^5) samples were drawn to generate mean ± std.
(C) Ten exemplary total GFP traces (dotted lines) and best fits (solid lines) of *elp6Δ* budding yeast cells and repressions r1 (left) and r2 (right).

(D) The median initial total GFP, GFP_0, is higher in *elp6Δ* traces better fitted by the repressor model (black) than in *elp6Δ* traces better fitted by the non-repressor model (red). This confirms that the repressor model fits induced *elp6Δ* cells better, while the non-repressor model fits uninduced *elp6Δ* cells. The number of cells and percentages of all GFP traces best fitted by the repressor model and non-repressor model are shown in brackets.
(E) Time to maximal mean total GFP is decreased in repression r2 for *elp6Δ* repressor cells (1.20 ± 0.14 *vs*. 0.69 ± 0.07). Bootstrap (10^5) samples of the repressor cells were drawn to generate mean ± std.
(F) Comparison of paired estimated single-cell parameters of *elp6Δ* cells of repression r1 and r2 show that median initial total GFP, GFP_0, median repression delay, t_{delay}, and median production rate, r_{prod}, are significantly different ($p = 1.2 \cdot 10^{-10}$, $p = 3.9 \cdot 10^{-5}$ and $p = 6.9 \cdot 10^{-8}$, respectively, two-sided paired sign test correcting for multiple testing with Bonferroni correction, m = 20, and the number of paired cells is 34), with both median GFP_0 and median r_{prod} increased and median t_{delay} decreased (median t_{delay} values of 0.95 and 0.50 h for r1 and r2, respectively) in r2.

elp6Δ shows stronger repression delay impairment in first repression

To first identify if the repression kinetics between wildtype and *elp6Δ* cells differed at the repressor subpopulation level, we compared the repression kinetics of wildtype and *elp6Δ* repressor cells. We found that the time to maximal mean total GFP increased for *elp6Δ* repressor cells for repression r1 (Figure 5A). Next, we wanted to identify if the *elp6Δ* repressor cells show altered repression kinetics at the single-cell level compared to wildtype cells. As we here assess different strains, we were no longer able to pair cells and, thus, compared the estimated single-cell parameters of all repressor cells. Using a two-sided median test and correcting for multiple testing, we discovered that the median t_{delay} for wildtype was significantly lower (median t_{delay} = 0.52h) compared to *elp6Δ* (median t_{delay} = 0.88h, $p = 4.7 \cdot 10^{-4}$) for r1 (Figure 5B). However, median GFP_0, median $r_{prod,}$ and median r_{deg} are comparable ($p = 1$, $p = 1$, and $p = 0.44$, respectively) (Figure 5B). This is in line with previous findings that wildtype and *elp6Δ* cells have similar Gal1 levels in induction i1 leading to comparable median GFP_0 in r1 and comparable production rates between wildtype and *elp6Δ* cells during induction i1 (*17*). Then, we performed the same analysis for repression r2. For r2, we found that the time to maximal mean total GFP at the repressor subpopulation level was more similar between wildtype and *elp6Δ* repressor cells (Figure 5C) than in r1. At the single-cell level, the median GFP_0, median t_{delay}, and median r_{prod} were significantly different ($p = 1.8 \cdot 10^{-6}$, $p = 2.2 \cdot 10^{-3}$, and $p = 8.1 \cdot 10^{-11}$, respectively) between wildtype and *elp6Δ* cells for r2 (Figure 5D) (see Materials and Methods), whereas median r_{deg} was comparable after correcting for multiple testing ($p = 0.04$) (see Materials and Methods). Overall, this analysis identifies a - compared to wildtype - stronger repression delay impairment of *elp6Δ* cells in r1, while the repression delay in r2 is more similar to wildtype cells. The previously described gain-of-induction-memory phenotype of *elp6Δ*, which showed identical Gal1 levels to wildtype in induction i1 but increased Gal1 levels in induction i2, is responsible for the increased median GFP_0 (*17*). Furthermore, Bheda et al. reported increased *elp6Δ* production rates for induction i2 but not for induction i1.

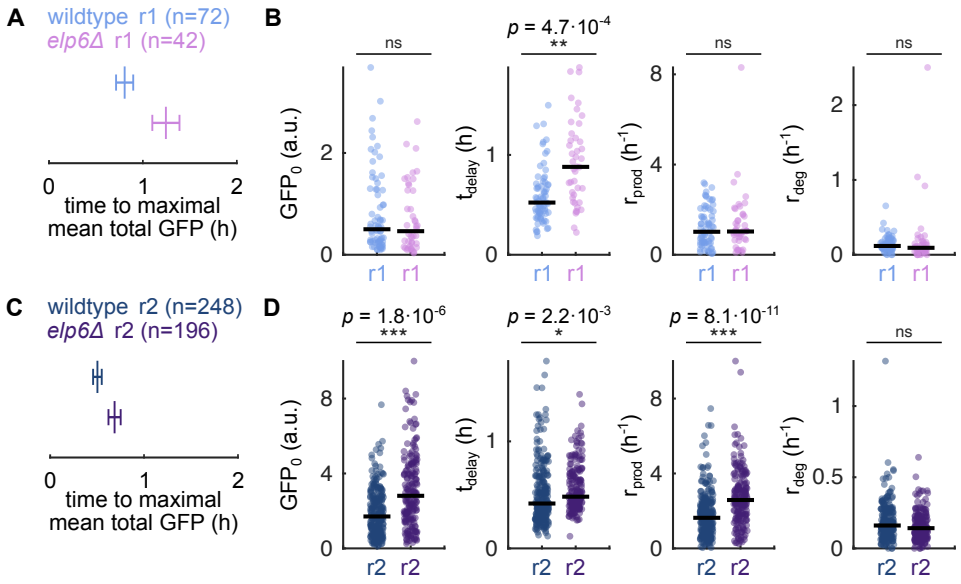


Figure 5. Stronger Gal1 repression delay impairment in *elp6Δ* cells during first repression.
(A) Time to maximal mean total GFP is increased for *elp6Δ* repressor cells for repression r1 (0.81 ± 0.09 *vs*. 1.20 ± 0.14). Bootstrap (10^5) samples were drawn to generate mean ± std.
(B) Comparison of estimated single-cell parameters of wildtype and *elp6Δ* cells of repression r1 shows that *elp6Δ* has a significantly different median repression delay, t_{delay}, in comparison to wildtype, ($p = 4.7 \cdot 10^{-4}$, two-sided median test correcting for multiple testing with Bonferroni correction, m = 20, and the number of cells for wildtype and *elp6Δ* is 72 and 42, respectively) with median t_{delay} increased in *elp6Δ* (median values of 0.52 and 0.88 h for wildtype and *elp6Δ*, respectively).
(C) Time to maximal mean total GFP is more comparable for wildtype and *elp6Δ* repressor cells for repression r2 (0.52 ± 0.05 *vs*. 0.69 ± 0.07). Bootstrap (10^5) samples were drawn to generate mean ± std.
(D) Comparison of estimated single-cell parameters of wildtype and *elp6Δ* cells of repression r2 shows median GFP_0, median t_{delay}, and median r_{prod}, to be significantly different in *elp6Δ* compared to wildtype ($p = 1.8 \cdot 10^{-6}$, $p = 2.2 \cdot 10^{-3}$, and $p = 8.1 \cdot 10^{-11}$, respectively, two-sided median test correcting for multiple testing with Bonferroni correction, m = 20 and number of cells for wildtype and *elp6Δ* is 248 and 196, respectively), with median GFP_0, median t_{delay}, and median r_{prod} increased in *elp6Δ* (median values of 0.42 and 0.48 h for wildtype and *elp6Δ*, respectively).

DISCUSSION

Based on population data, we found an earlier repression response upon a repeated repression. We used mathematical modeling to demonstrate that changes in the repression response were not simply due to different proportions of repressor cells. Interestingly, we discovered that *elp6Δ* cells showed prolonged repression delay in the first repression but a more comparable delay in comparison to wildtype in the second repression.

Repression memory manifests as faster initiation of second repression

At both the population and single-cell level, we identified that both wildtype and *elp6Δ* cells have an earlier repression response in repression r2 (Figures 3E–F, 4E–F, S1E-F, S2E-F). This implies that cells repeatedly exposed to glucose are faster at initiating Gal1 repression. We thus propose

that budding yeast cells do not only show reinduction memory when exposed to repeated galactose inductions (*17*) but also have a repression memory when exposed to repeated glucose repressions. We hypothesize that faster repression of the galactose-metabolizing machinery would save energy and/or allow for faster induction of the glucose-metabolizing machinery, which would be beneficial to individual yeast cells. It will be intriguing to investigate how Gal1 repression is acquired on a molecular level and if glucose and galactose repression memories are linked in the future. We would like to note here that we assessed the kinetics of the Gal1-GFP fusion protein, which has been successfully employed in several studies involving Gal1 gene expression (*16*, *23*–*26*), and hence also modeled Gal1-GFP kinetics. Moreover, it should be mentioned that potential photo-bleaching effects would not influence the estimated repression delays, and, thus, our repression memory hypothesis. When considering memory as the fold change of the second delay to the first delay, we discover reinduction memory to have an overall stronger effect size. The fold changes of the repression delays are (median t_{delay} r2/median t_{delay} r1 = 0.38/0.50 = 0.76) 0.76 for wildtype and (median t_{delay} r2/median t_{delay} r1 = 0.50/0.95 = 0.53) 0.53 for *elp6Δ*. In comparison, the fold changes of the induction delays are (median t_{delay} i2/median t_{delay} i1 = 0.67/2.33 = 0.29) 0.29 and (median t_{delay} i2/median t_{delay} i1 = 0.50/2.50 = 0.20) 0.20 for wildtype and elp6Δ, respectively, as reported in Bheda et al. (*17*).

elp6Δ is a novel gain-of-repression-memory mutant

We observed that *elp6Δ* cells had longer repression delay in repression r1 than wildtype cells, while the repression delay is more comparable for wildtype and *elp6Δ* cells in repression r2 at both the population and single-cell level (Figure 5). This reveals that the repression memory effects (shortening of repression delay in r2) of *elp6Δ* cells are stronger than in wildtype cells. Hence, *elp6Δ* can be classified as a gain-of-repression-memory mutant. *Elp6* has been identified as one of the six subunits of the so called RNA polymerase II elongator complex (*24*) and has been linked to a variety of biological functions (25–27). However, its precise function is still debated. Contrary to induction, where *elp6Δ* and wildtype show comparable induction delays in the first induction but exhibit significantly different induction delays in the second induction, *elp6Δ* and wildtype cells show significantly different repression delays in r1 and more comparable repression delays in r2 (*17*) (Figure 5B and D). Although reinduction memory has a stronger overall effect size than repression memory, the relative memory effect of *elp6Δ* and wildtype is comparable between induction (*elp6Δ*induction memory/wildtype induction memory = 0.20/0.29 = 0.69) and repression (*elp6Δ* repression memory/wildtype repression memory = 0.53/0.76 = 0.70). Whether the gain-of-repression memory is a unique characteristic of *elp6Δ* cells or if other knock-out mutants show similar repression kinetics and delay responses will be interesting to investigate in the future.

ACKNOWLEDGMENTS

L.S. was funded by the BMBF project TIDY (031L0170B). L.S. is especially grateful to the Technical University of Munich's Department of Mathematics, whose Entrepreneurial Award (within the Program "Global Challenges for Women in Math Science") contributed to the completion of this project. L.S. acknowledges further support by the Add-on Fellowship for Interdisciplinary Life Science of the Joachim Herz Foundation. C.M. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 866411). The work in R.S. laboratory was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through SFB 1064 (Project-ID 213249687) and SFB 1309 (Project-ID 325871075), the AmPro program (ZT0026) and Helmholtz Gesellschaft.

AUTHOR CONTRIBUTIONS

Conceptualization, P.B., R.S. and C.M.; Methodology, L.S.; Software, L.S.; Validation, L.S.; Formal Analysis, L.S.; Investigation, L.S., I.K.; Resources, R.S., and C.M.; Data Curation, L.S., M.B.; Writing – Original Draft, L.S.; Writing – Review & Editing, L.S., M.S., R.S. and C.M.; Visualization, L.S.; Supervision, P.B., R.S., and C.M.; Project Administration, R.S. and C.M.; Funding Acquisition, R.S., and C.M..

DECLARATION OF INTERESTS

The authors declare no conflict of interests.

MATERIALS AND METHODS

Data acquisition and sources

For the analysis for Figures 1-5, we used microscopy images and initial segmentation, mapping, and tracking information from a microfluidics experiment from Bheda et al. (*17*), which contained 13 and ten positions for wildtype and *elp6Δ*, respectively. The images from the first two hours of repression r1 were rectified, and the segmentation, mapping, and tracking were extended to the entire two hours of repression r2. Bheda et al. only segmented r2 partially since they were primarily interested in galactose induction, and did not adjust the final repression frames. Using the software PhyloCell (*19*), we manually corrected the segmentation, mapping, and tracking of r1 and r2 for wildtype and *elp6Δ*. For the replicate analysis (Figures S1 and S2), we repeated the induction-repression experiment as described in Bheda et al. (*17*) (Figure 2A). Due to low cell numbers, we pooled data from three and two independent experiments for wildtype and *elp6Δ*, totaling 13 and 23 positions, respectively. Using the software YeaZ (*20*) and Cell-ACDC (*21*) for cell segmentation, mapping and tracking, we extracted the relevant single-cell information of the live-cell images for both repressions r1 and r2. During glucose repression, the yeast cells proliferated, increasing the cell numbers within the microfluidic chambers. However, filled microfluidic chambers no longer assure that all the progeny of a cell is recorded, and mapping and tracking of cells become infeasible. To ensure mapping and tracking of single yeast cells within the microfluidic chambers, the glucose repressions were limited to a maximum of 4 h and the overall experiment was limited to 16 h (4 h in glucose (r0), 3 h in galactose (i1), 4 h in glucose (r1), 3 h in galactose (i2), 2 h in glucose (r2)).

Data preprocessing

We extracted the single-cell information relevant for our analysis, namely cell ID, mother cell ID, detection frame (first frame in which a cell is detected), last frame (last frame a cell is detected), relative GFP intensities per time (mean GFP intensity of a segmented cell) and cell area per time. As the data regarding the relative GFP intensities and cell area was not sorted over time, we first sorted it and then calculated the total GFP fluorescence per time given by

$$\text{total GFP fluorescence} = \text{relative GFP fluorescence} \times \text{cell area}.$$

Finally, cells that were not imaged till the end of the experiment, cells with missing relative GFP and/or cell area values, and cells that were supposedly detected before their mother cells (segmentation error) were discarded.

Dilution compensation

Cytoplasmic proteins are disseminated between the mother and daughter cells during budding. Assuming that Gal1 is not produced or degraded, protein redistribution causes a drop in total Gal1-GFP fluorescence in the mother cell and a rise in total Gal1-GFP fluorescence in the daughter cell till the mother and daughter cells split (Figure 2C top). As a result, regardless of repression, dilution causes variations in total Gal1-GFP fluorescence. The daughter cell grows to about ⅓ of the size of the mother cell (*27*) such that the decrease in total Gal1-GFP fluorescence

due to dilution was expected to be ⅓ of the initial total Gal1-GFP fluorescence of the mother cell. To ensure that dilution does not overshadow potentially more subtle repression kinetics, we created artificial non-dividing cells compensating for dilution by adding the total Gal1-GFP fluorescence of the progeny of a cell present at the start of glucose repression, which we called mother cell, to the total Gal1-GFP fluorescence of that mother cell during the first 2 h of repression (Figures 2C bottom, 2D, 4B, S1B, and S2B). For mother cells with a bud at the beginning of a repression period, we additionally added the bud to the total Gal1-GFP fluorescence of that mother cell. The GFP traces of all computed non-dividing cells can be found under https://github.com/marrlab/Gal1repression. As we found the maximal mean total GFP to be attained before 2 h of glucose exposure, we restricted our analysis to the first 2 h of repression.

Models

During the first two hours of glucose repression, we modeled the kinetics of the total GFP of every single cell. Due to the high variability in galactose induction, we assumed that our initial cell population at the beginning of repression contained induced and uninduced cells, which show and do not show repression kinetics, respectively. We developed two models, the repressor and the non-repressor model, to account for both total GFP kinetics during repression.

Non-repressor model

For more information regarding the non-repressor model, see the main text.

Repressor model

For more information regarding the repressor model, see the main text.

Noise model

Experimental data, such as total GFP per cell per time, is noise corrupted. As a result, we used an underlying additive Gaussian noise model with a constant variance σ^2 throughout time to test our models. The single-cell specific model parameters are comprised in the parameter vector Θ_i for cell i and the experimental measurement at time point k for cell i is denoted by $\underline{y_i}^k$. The log-likelihood for the Gaussian noise model is given by

$$\log L(\Theta_i) = -\tfrac{1}{2} \sum_k \left(\log(2\pi\sigma_i^2)+(\underline{y_i}^k - y(t_k,\Theta_i))^2/\sigma_i^2\right).$$

We obtained the optimal model parameters of both models for the total GFP traces for each cell by performing maximum likelihood estimation.

Optimization and parameter estimation

For each total GFP trace separately for r1 and r2, wildtype, and *elp6Δ*, we computed the model parameters for both the non-repressor and repressor models. The initial total GFP GFP_0, the basal production rate r_{basal}, and the degradation rate r_{deg} are the model parameters for the non-repressor model. Instead of a basal production rate, r_{basal}, we have a production rate, r_{prod}, for the

repressor model. Also, we discovered the time point of delayed repression t_{delay} for the repressor model. For both models we also infer one noise parameter σ determining the spread of the Gaussian noise model. We assumed that all parameters are constant over time. For numerical reasons we optimized the parameters in $\log_{10}$ scale (*28*) and rescaled the data by 10^7. As total Gal1-GFP fluorescence signal and total Gal1-GFP molecules are (linearly) mapped by an unknown constant, the number of total Gal1-GFP molecules is always scalable by that unknown constant that we exploit to increase convergence. The lower and upper bounds for all initial, rate, and noise parameters are –10 and 1 in $\log_{10}$ scale, respectively, assuring that the whole range of biologically plausible parameter values is covered. The lower and upper bounds for the repression delay are given by 36 s and 2 h (corresponding to –2 and $\log_{10}(2)$ in $\log_{10}$ scale). As we only considered 2 h of glucose repression, we did not allow the time delay to take on larger values. We performed multi-start local optimization of the negative log-likelihood using the parameter estimation toolbox PESTO (*29*). For each model and total GFP trace, we performed local optimization runs from at least 20 different Latin-hypercube-sampled starts. If less than five starts converged, i.e. the objective function values of the starts differ less than 0.1 to the best start, we re-ran the optimization with 50, 100, and 200 starts until at least five starts converged for each GFP trace.

Model selection

We used the Bayesian information criterion (BIC) (*30*) for comparing the non-repressor model and the repressor model per total GFP trace. The BIC is calculated by

$$BIC = \log(n)k - 2\log L,$$

where n is the number of data points, k is the number of estimated parameters and logL is the log-likelihood value for the maximum likelihood estimate of the model parameters. Here, the number of estimated parameters is either four for the non-repressor model or five for the repressor model. The BIC rewards high likelihood values and penalizes the model complexity in the form of additional model parameters. We considered the repressor model to fit a given total GFP trace considerably better than the non-repressor model if $BIC_{repressor} < BIC_{non\text{-}repressor} - 10$ (Figures 3C–D, 4C–D, S1C-D and S2C-D).

Statistical analysis

Comparison of initial total GFP of total GFP traces

On the estimated initial total GFP, GFP_0, of all total GFP traces significantly better fitted by a repressor model and all total GFP traces better fitted by a non-repressor model, we did a two-sided median test. To avoid false-positive results, we used the Bonferroni correction, which adjusts the significance-level $\alpha = 0.05$ by the total number of investigated null hypotheses m, such that $\alpha' = \alpha/m$. In this study, the total number of null hypotheses for the main analysis is m = 20:

- Two hypothesis tests comparing initial total GFP between repressor and non-repressor fits for wildtype r1 and r2,

- Four hypothesis tests comparing estimated single-cell parameters between wildtype r1 and r2,
- Two hypothesis tests comparing initial total GFP between repressor and non-repressor fits for *elp6Δ* r1 and r2,
- Four hypothesis tests comparing estimated single-cell parameters between *elp6Δ* r1 and r2,
- Four hypothesis tests comparing estimated single-cell parameters between wildtype and *elp6Δ* for r1, and
- Four hypothesis tests comparing estimated single-cell parameters between wildtype and *elp6Δ* for r2.

In this study, the total number of null hypotheses for the replicate analysis is m = 12:

- Two hypothesis tests comparing initial total GFP between repressor and non-repressor fits for wildtype r1 and r2,
- Four hypothesis tests comparing estimated single-cell parameters between wildtype r1 and r2,
- Two hypothesis tests comparing initial total GFP between repressor and non-repressor fits for *elp6Δ* r1 and r2, and
- Four hypothesis tests comparing estimated single-cell parameters between *elp6Δ* r1 and r2.

Comparison of estimated single-cell parameters between repression r1 and r2 for wildtype and elp6Δ

We ignored all traces that were well described by a non-repressor model (uninduced cells) and focused the statistical analysis on the total GFP traces for which the repressor model gave a considerably better fit (see the Model selection, Figures 3C, 4C, S1C and S2C). For those total GFP traces, we compared the estimated single-cell parameters of initial total GFP, GFP_0, repression delay t_{delay}, production and degradation rates, r_{prod} and r_{deg}, for repressions r1 and r2 and wildtype and *elp6Δ*. We performed a two-sided paired sign test on the estimated single-cell parameters of paired mother cells in r1 and r2 for both wildtype (Figures 3F and S1F) and *elp6Δ* (Figures 4F and S2F). To avoid false-positive outcomes, we used the Bonferroni correction, which modified the significance level of $\alpha = 0.05$ by the total number of tested null hypotheses m to $\alpha' = \alpha/m$, with m = 20 for the main analysis and m = 12 for the replicate analysis.

Comparison of estimated single-cell parameters between wildtype and elp6Δ for repression r1 and r2

We performed a two-sided median test on the estimated single-cell parameters of all cells for wildtype and *elp6Δ* for r1 (Figure 5B) and r2 (Figure 5D) of the main analysis. Due to the comparison between two different yeast strains, we could not perform a paired test. To counteract false-positive results, we again corrected for multiple testing according to the Bonferroni correction, where the significance-level $\alpha = 0.05$ was adjusted by the total number of tested null hypotheses m to $\alpha' = \alpha/m$ with m = 20 for the main analysis. As the replicate analysis was

performed on pooled data to increase overall cell numbers, we were not able to compare the estimated single-cell parameters between wildtype and *elp6Δ*.

Implementation

The toolboxes used for segmentation, mapping, and tracking are available at https://github.com/gcharvin/phyloCell (PhyloCell (*19*)), https://github.com/lpbsscientist/YeaZ-GUI (YeaZ (*20*)), and https://github.com/SchmollerLab/Cell_ACDC (Cell-ACDC (*21*)). The toolbox used for parameter estimation (PESTO (*29*)) is available under https://github.com/ICB-DCM. The MATLAB code corresponding to this manuscript is available under https://github.com/marrlab/Gal1repression. The analysis was performed with MATLAB 2017b.

REFERENCES

1. M. Acar, A. Becskei, A. van Oudenaarden, Enhancement of cellular memory by reducing stochastic transitions. *Nature*. **435**, 228–232 (2005).
2. K. Liu, Y. Li, V. Prabhu, L. Young, K. G. Becker, P. J. Munson, N.-P. Weng, Augmentation in Expression of Activation-Induced Genes Differentiates Memory from Naive CD4 T Cells and Is a Molecular Mechanism for Enhanced Cellular Response of Memory CD4 T Cells. *The Journal of Immunology*. **166** (2001), pp. 7335–7344.
3. Y. Araki, Z. Wang, C. Zang, W. H. Wood 3rd, D. Schones, K. Cui, T.-Y. Roh, B. Lhotsky, R. P. Wersto, W. Peng, K. G. Becker, K. Zhao, N.-P. Weng, Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. *Immunity*. **30**, 912–925 (2009).
4. C. V. Mobbs, Glucose-Induced Transcriptional Hysteresis: Role in Obesity, Metabolic Memory, Diabetes, and Aging. *Front. Endocrinol.* . **9**, 232 (2018).
5. L. M. Villeneuve, M. A. Reddy, R. Natarajan, Epigenetics: deciphering its role in diabetes and its chronic complications. *Clin. Exp. Pharmacol. Physiol.* **38**, 451–459 (2011).
6. C. V. Mobbs, Genetic influences on glucose neurotoxicity, aging, and diabetes: a possible role for glucose hysteresis. *Genetica*. **91**, 239–253 (1993).
7. A. J. Courey, S. Jia, Transcriptional repression: the long and the short of it. *Genes Dev.* **15**, 2786–2796 (2001).
8. E. Maldonado, M. Hampsey, D. Reinberg, Repression: targeting the heart of the matter. *Cell*. **99**, 455–458 (1999).
9. H. W. Kosterlitz, The fermentation of galactose and galactose-1-phosphate. *Biochemical Journal*. **37** (1943), pp. 322–326.
10. L. F. Leloir, The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. *Archives of Biochemistry and Biophysics*. **33** (1951), pp. 186–190.
11. H. C. Douglas, D. C. Hawthorne, Enzymatic expression and genetic linkage of genes controlling galactose utilization in Saccharomyces. *Genetics*. **49** (1964), pp. 837–844.
12. D. G. Brickner, I. Cajigas, Y. Fondufe-Mittendorf, S. Ahmed, P.-C. Lee, J. Widom, J. H. Brickner, H2A.Z-Mediated Localization of Genes at the Nuclear Periphery Confers Epigenetic Memory of Previous Transcriptional State. *PLoS Biology*. **5** (2007), p. e81.
13. S. Kundu, P. J. Horn, C. L. Peterson, SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. *Genes & Development*. **21** (2007), pp. 997–1004.
14. V. Sood, I. Cajigas, A. D'Urso, W. H. Light, J. H. Brickner, Epigenetic Transcriptional Memory of GAL Genes Depends on Growth in Glucose and the Tup1 Transcription Factor in Saccharomyces cerevisiae. *Genetics*. **206** (2017), pp. 1895–1907.
15. S. R. Stockwell, C. R. Landry, S. A. Rifkin, The yeast galactose network as a quantitative model for cellular memory. *Mol. Biosyst.* **11**, 28–37 (2015).
16. I. Zacharioudakis, T. Gligoris, D. Tzamarias, A yeast catabolic enzyme controls transcriptional memory. *Curr. Biol.* **17**, 2041–2046 (2007).

17. P. Bheda, D. Aguilar-Gómez, N. B. Becker, J. Becker, E. Stavrou, I. Kukhtevich, T. Höfer, S. Maerkl, G. Charvin, C. Marr, A. Kirmizis, R. Schneider, Single-Cell Tracing Dissects Regulation of Maintenance and Inheritance of Transcriptional Reinduction Memory. *Mol. Cell*. **78**, 915–925.e7 (2020).
18. B. B. Lee, A. Choi, J. H. Kim, Y. Jun, H. Woo, S. D. Ha, C. Y. Yoon, J.-T. Hwang, L. Steinmetz, S. Buratowski, S. Lee, H. Y. Kim, T. Kim, Rpd3L HDAC links H3K4me3 to transcriptional repression memory. *Nucleic Acids Res.* **46**, 8261–8274 (2018).
19. Y. Goulev, S. Morlot, A. Matifas, B. Huang, M. Molin, M. B. Toledano, G. Charvin, Nonlinear feedback drives homeostatic plasticity in HO stress response. *Elife*. **6** (2017), doi:10.7554/eLife.23971.
20. N. Dietler, M. Minder, V. Gligorovski, A. M. Economou, D. A. H. L. Joly, A. Sadeghi, C. H. M. Chan, M. Koziński, M. Weigert, A.-F. Bitbol, S. J. Rahi, A convolutional neural network segments yeast microscopy images with high accuracy. *Nat. Commun.* **11**, 5723 (2020).
21. F. Padovani, B. Mairhörmann, P. Falter-Braun, J. Lengefeld, K. M. Schmoller, Cell-ACDC: a user-friendly toolset embedding state-of-the-art neural networks for segmentation, tracking and cell cycle annotations of live-cell imaging data, *bioRxiv 2021.09.28.462199*, doi:10.1101/2021.09.28.462199.
22. M. Johnston, A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiological Reviews*. **51** (1987), pp. 458–476.
23. J. Li, S. Wang, W. J. VanDusen, L. D. Schultz, H. A. George, W. K. Herber, H. J. Chae, W. E. Bentley, G. Rao, Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter. *Biotechnol. Bioeng.* **70**, 187–196 (2000).
24. G. I. Lang, D. Botstein, A test of the coordinated expression hypothesis for the origin and maintenance of the GAL cluster in yeast. *PLoS One*. **6**, e25290 (2011).
25. R. K. Niedenthal, L. Riles, M. Johnston, J. H. Hegemann, Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast*. **12**, 773–786 (1996).
26. S. R. Biggar, Cell signaling can direct either binary or graded transcriptional responses. *The EMBO Journal*. **20** (2001), pp. 3167–3176.
27. I. V. Kukhtevich, N. Lohrberg, F. Padovani, R. Schneider, K. M. Schmoller, Cell size sets the diameter of the budding yeast contractile ring. *Nat. Commun.* **11**, 2952 (2020).
28. H. Hass, C. Loos, E. Raimúndez-Álvarez, J. Timmer, J. Hasenauer, C. Kreutz, Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics*. **35** (2019), pp. 3073–3082.
29. P. Stapor, D. Weindl, B. Ballnus, S. Hug, C. Loos, A. Fiedler, S. Krause, S. Hroß, F. Fröhlich, J. Hasenauer, PESTO: Parameter EStimation TOolbox. *Bioinformatics*. **34**, 705–707 (2018).
30. G. Schwarz, Estimating the Dimension of a Model. *The Annals of Statistics*. **6** (1978), , doi:10.1214/aos/1176344136.

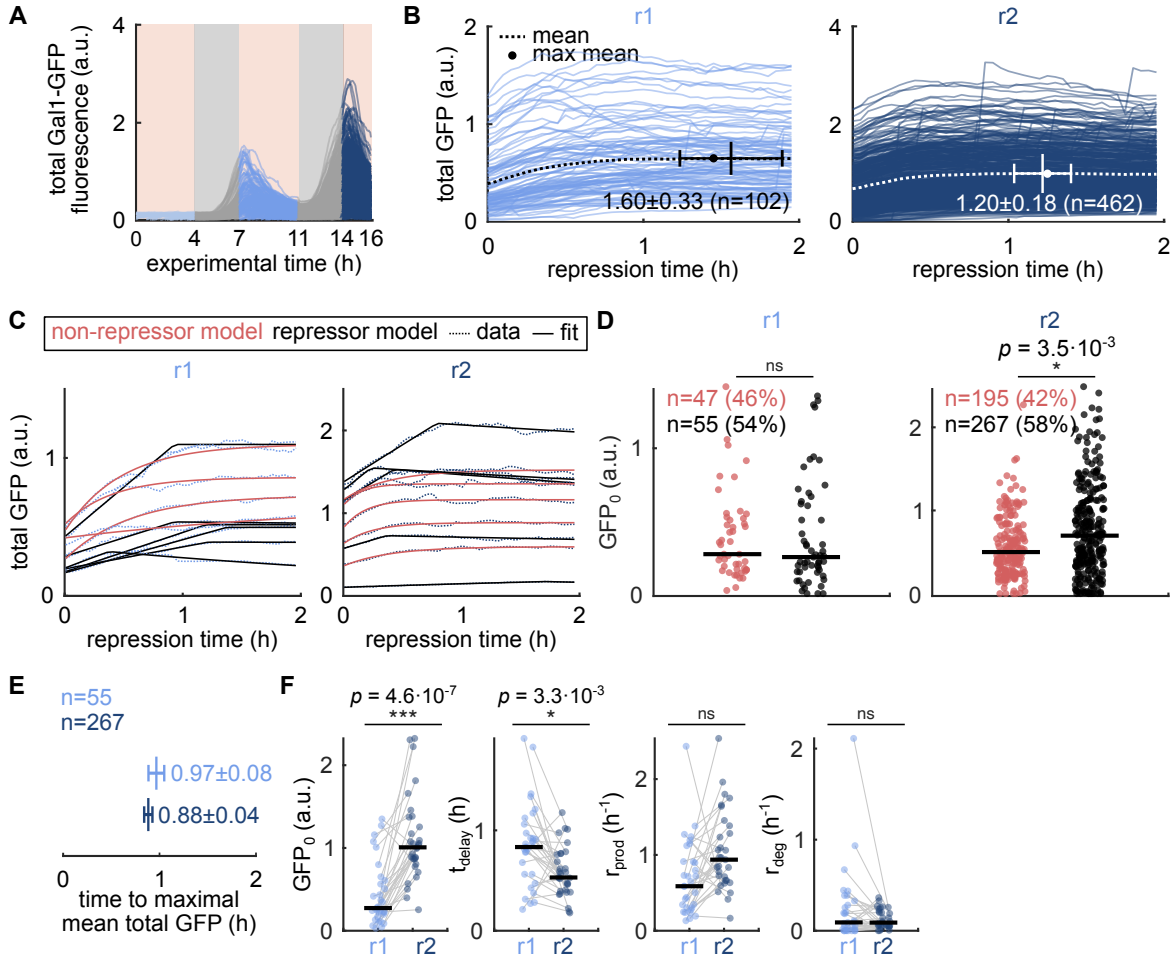


Figure S1. Earlier repression response in repression r2 at both the population and single-cell level for cells of replicate experiment.

(A) Single-cell traces of total Gal1-GFP fluorescence signal across two inductions i1 and i2 (gray) and repressions r0, r1, and r2 (blue).

(B) Single-cell traces of total GFP signal adjusted for dilution (see (C)) for the first two hours of repression r1 (left) and repression r2 (right). Time to maximal mean total GFP is 24 min shorter in repression r2, where mean total GFP is indicated by the dotted line and the maximal mean total GFP is highlighted by the dot. Bootstrap (10^5) samples were drawn to generate mean ± std.

(C) Ten exemplary total GFP traces (dotted lines) and best fits (solid lines) for repressions r1 (left) and r2 (right). GFP traces best fitted with a repressor model are shown in black and fits of total GFP traces best fitted with a non-repressor model are shown in red.

(D) The median initial total GFP, GFP_0, is higher in traces better fitted by the repressor model (black) than in traces better fitted by the non-repressor model (red). This confirms that the repressor model fits induced cells better, while the non-repressor model fits uninduced cells. The number of cells and percentages of all GFP traces best fitted by the repressor model and non-repressor model are shown.

(E) Time to maximal mean total GFP is decreased in repression r2 for repressor cells (0.97 ± 0.08 vs. 0.88 ± 0.04). Bootstrap (10^5) samples of the repressor cells were drawn to generate mean ± std.

(F) Comparison of paired estimated single-cell parameters of repression r1 and r2 shows that the median initial total GFP, GFP_0, and median repression delay, t_{delay}, are significantly different ($p = 4.6 \cdot 10^{-7}$ and $p = 3.3 \cdot 10^{-3}$, respectively, two-sided paired sign test correcting for multiple testing with Bonferroni correction, m = 12, and the number of paired cells = 31), with median GFP_0 increased and median t_{delay} decreased (median t_{delay} values of 0.83 and 0.53 for r1 and r2, respectively).

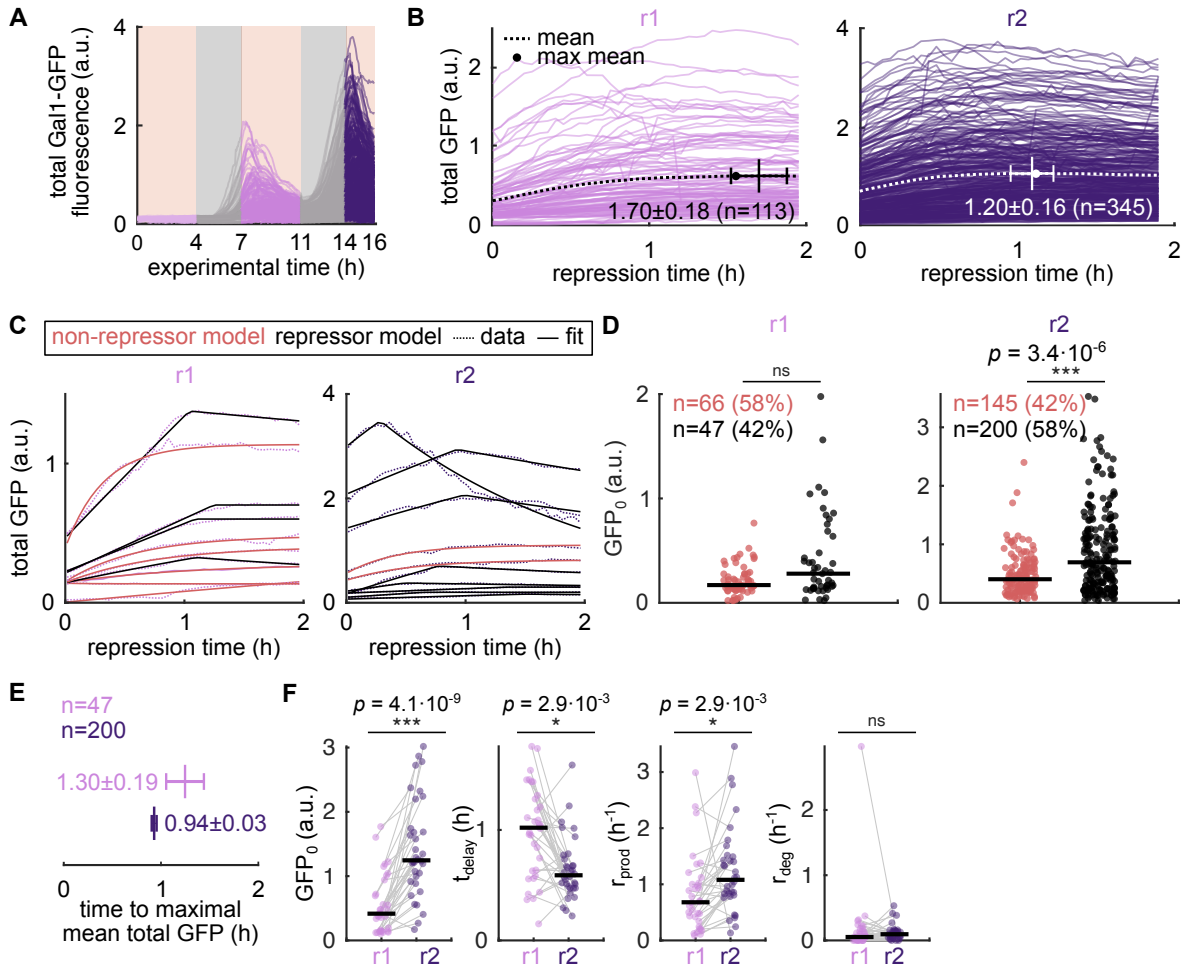


Figure S2. Earlier repression response in repression r2 at both the population and single-cell level for *elp6Δ* cells of replicate experiment.

(A) Single-cell traces of total Gal1-GFP fluorescence signal of *elp6Δ* budding yeast cells across two inductions i1 and i2 (gray) and repressions 0, 1, and 2 (purple).

(B) Single-cell traces of total GFP signal of *elp6Δ* budding yeast cells adjusted for dilution (see Figure 2C) for the first two hours of r1 (left) and r2 (right). Time to maximal mean total GFP is 36 min shorter in repression r2, where the mean total GFP signal is indicated by the dotted line and the maximal mean total GFP is highlighted by the dot. Bootstrap (10^5) samples were drawn to generate mean ± std.

(C) Ten exemplary total GFP traces (dotted lines) and best fits (solid lines) of *elp6Δ* budding yeast cells and repressions r1 (left) and r2 (right).

(D) The median initial total GFP, GFP_0, is higher in *elp6Δ* traces better fitted by the repressor model (black) than in *elp6Δ* traces better fitted by the non-repressor model (red). This confirms that the repressor model fits induced *elp6Δ* cells better, while the non-repressor model fits uninduced *elp6Δ* cells. The number of cells and percentages of all GFP traces best fitted by the repressor model and non-repressor model are shown.

(E) Time to maximal mean total GFP is decreased in repression r2 for *elp6Δ* repressor cells (1.30 ± 0.19 *vs*. 0.94 ± 0.02). Bootstrap (10^5) samples of the repressor cells were drawn to generate mean ± std.

(F) Comparison of paired estimated single-cell parameters of *elp6Δ* cells of repression r1 and r2 show that median initial total GFP, GFP_0, median repression delay, t_{delay}, and median production rate, r_{prod}, are significantly different ($p = 4.1 \cdot 10^{-9}$, $p = 2.9 \cdot 10^{-3}$, and $p = 2.9 \cdot 10^{-3}$, respectively, two-sided paired sign test correcting for multiple testing with Bonferroni correction, m = 12, and the number of paired cells is 34), with median GFP_0 and r_{prod} increased and median t_{delay} decreased (median values of 1.00 and 0.59 h for r1 and r2, respectively) in r2.