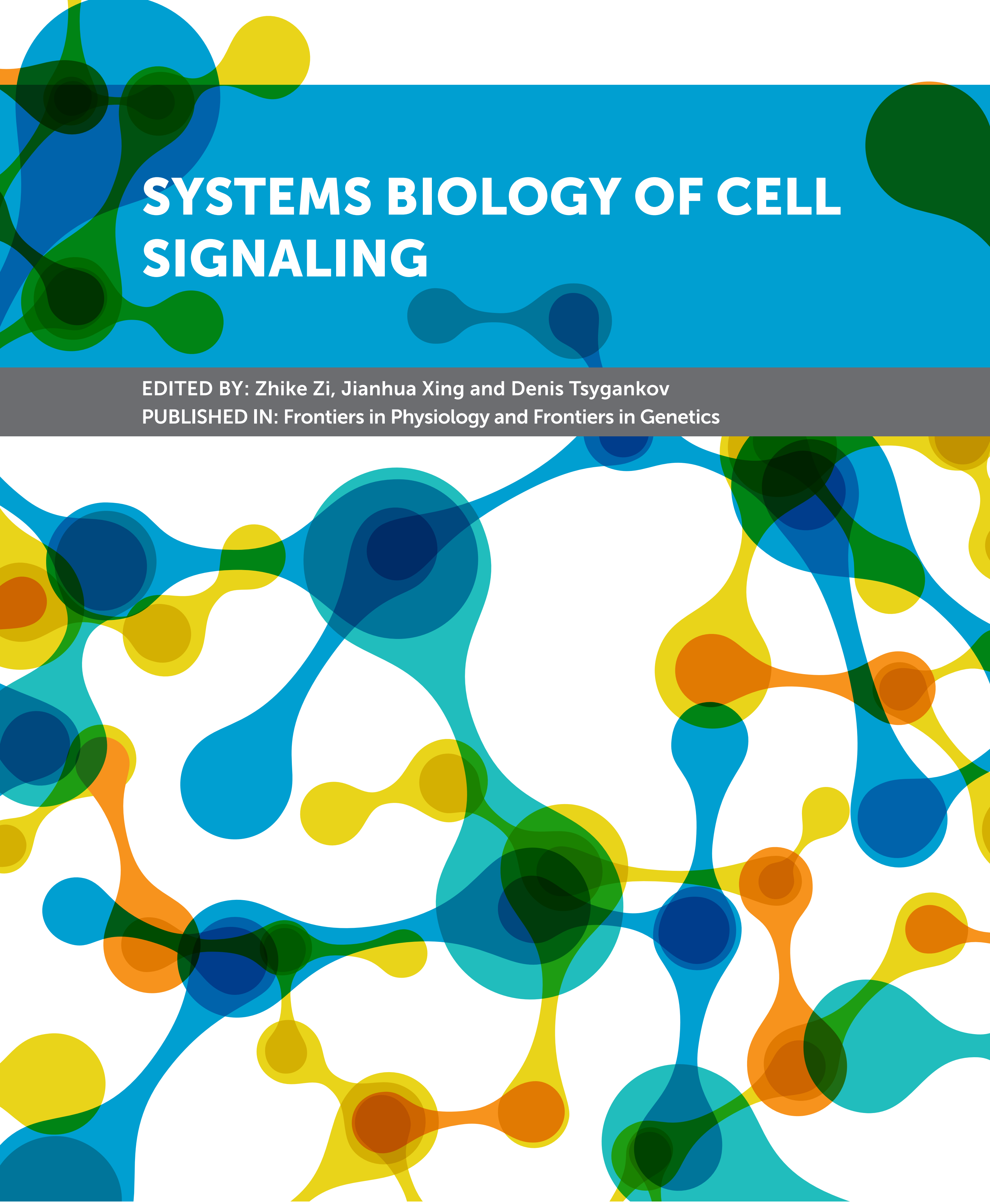


SYSTEMS BIOLOGY OF CELL SIGNALING

EDITED BY: Zhike Zi, Jianhua Xing and Denis Tsygankov

PUBLISHED IN: Frontiers in Physiology and Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-281-3
DOI 10.3389/978-2-88974-281-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SYSTEMS BIOLOGY OF CELL SIGNALING

Topic Editors:
Zhike Zi, Max Planck Institute for Molecular Genetics, Germany
Jianhua Xing, University of Pittsburgh, United States
Denis Tsygankov, Georgia Institute of Technology, United States

Topic Editor Prof. Xing is in collaboration with ATCC (https://www.atcc.org/) on testing some of their cell lines in research. All other Topic Editors declare no competing interests with regards to the Research Topic subject.

Citation: Zi, Z., Xing, J., Tsygankov, D., eds. (2022). Systems Biology of Cell Signaling. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-281-3

Table of Contents

frontiers in Physiology

MINI REVIEW
published: 16 April 2020
doi: 10.3389/fphys.2020.00327

A Systems Biology Approach Identifies Hidden Regulatory Connections Between the Circadian and Cell-Cycle Checkpoints

Xianlin Zou[1], Dae Wook Kim[2], Tetsuya Gotoh[1†], Jingjing Liu[1†], Jae Kyoung Kim[2] and Carla V. Finkielstein[1]*

[1] Integrated Cellular Responses Laboratory, Department of Biological Sciences, Fralin Life Sciences Institute, Virginia Tech, Blacksburg, VA, United States, [2] Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

OPEN ACCESS

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
Alexander Loewer,
Darmstadt University of Technology, Germany
Didier Gonze,
Université libre de Bruxelles, Belgium

***Correspondence:**
Carla V. Finkielstein
finkielc@vt.edu

†Present address:
Tetsuya Gotoh,
Department of Research and Development, WAKO-RIKEN Institute, Protects Co., Ltd., Saitama, Japan
Jingjing Liu,
Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, United States

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *03 February 2020*
Accepted: *20 March 2020*
Published: *16 April 2020*

Citation:
Zou X, Kim DW, Gotoh T, Liu J, Kim JK and Finkielstein CV (2020) A Systems Biology Approach Identifies Hidden Regulatory Connections Between the Circadian and Cell-Cycle Checkpoints. Front. Physiol. 11:327. doi: 10.3389/fphys.2020.00327

Circadian rhythms form a self-sustaining, endogenous, time-keeping system that allows organisms to anticipate daily environmental changes. The core of the clock network consists of interlocking transcriptional-translational feedback loops that ensures that metabolic, behavioral, and physiological processes run on a 24 h timescale. The hierarchical nature of the clock manifests itself in multiple points of control on the daily cell division cycle, which relies on synthesis, degradation, and post-translational modification for progression. This relationship is particularly important for understanding the role of clock components in sensing stress conditions and triggering checkpoint signals that stop cell cycle progression. A case in point is the interplay among the circadian factor PERIOD2 (PER2), the tumor suppressor p53, and the oncogenic mouse double minute-2 homolog protein (MDM2), which is the p53's negative regulator. Under unstressed conditions, PER2 and p53 form a stable complex in the cytosol and, along with MDM2, a trimeric complex in the nucleus. Association of PER2 to the C-terminus end of p53 prevents MDM2-mediated ubiquitylation and degradation of p53 as well as p53's transcriptional activation. Remarkably, when not bound to p53, PER2 acts as substrate for the E3-ligase activity of MDM2; thus, PER2 is degraded in a phosphorylation-independent fashion. Unexpectedly, the phase relationship between PER2 and p53 are opposite; however, a systematic modeling approach, inferred from the oscillatory time course data of PER2 and p53, aided in identifying additional regulatory scenarios that explained, *a priori*, seemingly conflicting experimental data. Therefore, we advocate for a combined experimental/mathematical approach to elucidating multilevel regulatory cellular processes.

Keywords: circadian rhythms, tumor suppressor, checkpoint signaling, clock genes, p53, mathematical modeling, systematic approach

INTRODUCTION

Circadian oscillators provide living organisms an adaptive advantage by enabling them to anticipate the demands of an evolving environment. To accomplish this, organisms synchronize their metabolism and behavior to external cues through a core molecular mechanism, named the circadian clock, that is driven by interlocking transcriptional-translational feedback loops and is linked to regulatory output pathways (Bell-Pedersen et al., 2005). Briefly, the mammalian core clock system is sustained by three interlocking mechanisms that are controlled by a combination of timely distributed factors whose binding and dissociation to the gene's regulatory regions dictate its phase of expression (for review see Takahashi, 2017 and references within). At the center of the network is the heterodimeric complex formed by the basic helix-loop-helix (bHLH)-PAR-ARNT-SIM (PAS) transcription factors CLOCK (circadian locomotor output cycles kaput and its paralog the Neuronal PAS domain Protein 2 NPAS2) and BMAL1 (Brain and Muscle ARNT-like 1 or the Aryl hydrocarbon Receptor Nuclear Translocator-Like protein 1 ARNTL). The CLOCK:BMAL1 complex binds to CAC(G/A)TG response elements (E-box, enhancer box) and activates the expression of *PERIOD* 1-3 (*PER1*, *PER2*, *PER3*) and *CRYPTOCHROME* 1-2 (*CRY1*, *CRY2*) genes. PER and CRY proteins accumulate asymmetrically during the day in the cytosol. PERs are targeted by casein kinases 1δ/ε (CK1δ/ε), phosphorylated, and degraded by the proteasome until the subsequent accumulation of CRY, later in the day, favors the formation of the PER:CRY:CK1δ/ε complex. This complex then shuttles to the nucleus where it interacts with CLOCK:BMAL1 to repress the transcription of *PER* and *CRY* genes. As *de novo* synthesis of PER and CRY molecules falls and the PER:CRY:CK1δ/ε complex is degraded, repression is released and CLOCK:BMAL1 becomes available for a new round of transcription. The CLOCK:BMAL1 complex also controls a second regulatory loop as it activates the expression of the *NR1D1* and *NR1D2* genes encoding the nuclear receptors REV-ERBα and β, respectively. The REV proteins compete with RORα/β/γ for RORE binding elements within the *BMAL1* promoter leading to repression or activation of *BMAL1*, respectively, which is antiphase to that of the *PER* genes. In a third regulatory loop, CLOCK:BMAL1 transcriptionally activates the expression of *DBP* (D-box binding PAR bZIP transcription factor) whose protein product binds to D-box response elements found in the regulatory regions of core clock components. Regardless, we have departed from the long-held belief that global rhythms of mRNA expression were exclusively driven by *de novo* rhythms in transcription and have embraced the contribution of post-transcriptional regulatory mechanisms in generating cycling messengers (Koike et al., 2012).

Coupling Between the Circadian Clock and the Cell Division Cycle

Like the self-sustained and cell-autonomous circadian clock, mammalian cells undergo periodic oscillations in the form of successive rounds of division. Proliferative cells progress through the cell cycle in a process whose completion takes about a day. In a manner resembling the circadian clock, phase transitions, from Gs to S and M and back in the cell cycle, are driven by *de novo* transcription, protein accumulation, post-translational modifications, and protein degradation. As a result, the existence of a coupling mechanism between the two oscillators was initially described (Brown, 1991; Bjarnason et al., 1999, 2001) and later proven to exist in organisms as diverse as the prokaryotic cyanobacterium *Synechococcus* PCC 7942 (Mori et al., 1996; Yang et al., 2010), the fungi *Neurospora crassa* (Hong et al., 2014), zebrafish (Tamai et al., 2012), and mammalian cells (Matsuo et al., 2003; Nagoshi et al., 2004; Kowalska et al., 2013). Later work further refined the circadian-cell cycle relationship to establish temporal windows in the circadian cycle at which specific cell cycle transitions were more likely to occur (for review see Gaucher et al., 2018). This phenomenon, usually referred to as "*circadian gating of the cell cycle*," has been proposed to provide a fitness benefit by, for example, ensuring that DNA replication takes place at times in which genotoxic stress is at minimum and metabolic conditions favor low levels of free radicals (Destici et al., 2011).

Of the many players involved in cell cycle progression, several key regulatory factors are directly controlled by circadian proteins. In a landmark study by Matsuo et al. (2003), the authors established that CLOCK:BMAL1 regulates the oscillatory expression of *Wee1*, a gene whose kinase product modulates the activity of cyclin B/Cdc2 and, therefore, G2/M progression in regenerating mouse liver. Accordingly, this mitotic clock-dependent gate was found to be disrupted in *cry1* deficient mice (Matsuo et al., 2003). Circadian factors also exert control over G1/S progression by transcriptionally modulating *Cdkn1a*, a gene that encodes for the cyclin/Cdk inhibitor p21$^{\mathrm{WAF1/CIP1}}$ (Grechez-Cassiau et al., 2008). Biochemical data backs the findings that clock orphan nuclear receptors REV-ERBα/β and RORα/γ exhibit antagonistic activities over *Cdkn1a* whereas genetic studies establish that their control is mediated by *Bmal1* (Grechez-Cassiau et al., 2008). A different G1 inhibitor, p16$^{\mathrm{ink4a}}$, which binds Cdk4/6 when it is dissociated from their cyclin counterpart, is also targeted for circadian regulation in primary fibroblasts (Kowalska et al., 2013). Accordingly, the complex between the nuclear RNA-binding protein NONO (non-POU domain-containing octamer-binding protein) and the circadian factor PER1 (NONO:PER1) drives the expression of *Cdkn2a* and, thus, the production of p16$^{\mathrm{ink4a}}$ in a manner that resembles PER1 expression (Kowalska et al., 2013). In fact, downregulation of NONO resulted in a non-rhythmic, low level expression of *Cdkn2a* and deregulation of G1 progression at a specific circadian phase (Kowalska et al., 2013). An additional key player in the process of cell growth and proliferation, the *c-Myc* oncogene is post-translationally regulated by CRY2 and the SCF$^{\mathrm{FBXL3}}$ ubiquitin ligase complex (Huber et al., 2016). In this scenario, CRY2-dependent turnover of MYC relies on MYC's phosphorylation and its recognition by SCF$^{\mathrm{FBXL3}}$ in a process in which CRY2 seems to act as an adaptor/chaperone/presenter molecule for the E3 ubiquitin complex (Huber et al., 2016). Studies in more complex and heterogenous multicellular system, i.e., enteroids, established circadian gating of cell division is mediated by intercellular signals arising from differentiated cells (Matsu-Ura et al., 2016). Consequently, multi-level regulatory mechanisms seem to ensure that gating occurs at specific windows of time in a tightly regulated process.

Checkpoint mechanisms ensure the faithful completion of each discrete phase of the cell cycle before the next one proceeds. Checkpoints monitor, for example, that DNA duplication is completed, that accurate chromosome segregation occurred, and that the integrity of the genome remains intact.

When considering the relevance of these processes in the context of "*timely progression*" through each cell cycle phase, it seems reasonable to speculate that clock mechanisms and checkpoint pathways would likely intersect. Today, evidence shows that crosstalk exists and occurs at multiple signaling levels; however, few studies explored the functional consequences beyond a descriptive level of analysis. For example, when PER1 is ectopically expressed, the checkpoint kinases Chk2 and ataxia-telangiectasia mutated (ATM) were found to co-immunoprecipitate with PER1 (Gery et al., 2006). Similarly, overexpression of PER1downregulated the levels of *Wee1*, *CcnB1* (encodes cyclin B1), and *Cdk1* (encodes Cdc2) mRNA and suppressed growth in cultured cancer cells (Gery et al., 2006). Another example is the product of human *Timeless* (hTIM), a gene with a well-established function in the *Drosophila* clock system (dTim) but with a debatable circadian role in mammalian cells (Mazzoccoli et al., 2016). Indeed, hTIM has little functional or structural resemblance to the dTim ortholog involved in clock function. It has a closer connection to dTim2, a molecule with a yet unclear role in circadian rhythms but which has relevance in DNA metabolism and chromosomal stability and integrity (Mazzoccoli et al., 2016). In this context, hTIM has been shown to bind to CRY2 and, in response to damage and replication stress, to CHK1 (Unsal-Kacmaz et al., 2005). Undoubtedly, one of the best-established connections between circadian components and checkpoint signaling arises from the findings that CRY1 and CRY2 have distinct roles in maintaining the integrity of the genome in response to genotoxic stress (Papp et al., 2015). In mammalian cells, CRY1 is phosphorylated and deubiquitylated by the Herpes virus associated ubiquitin-specific protease (HAUSP/Usp7) while CRY2 is destabilized through its interaction with SCF^{FBXL3} in response to DNA-damage (Papp et al., 2015). This counter effect on CRY proteins tilt their expression balance and, therefore, influences the targeted transcription of downstream genes in a manner that protects genomic stability (Papp et al., 2015).

More recently, we reported that the circadian factor PER2 directly binds to the tumor suppressor and G1/S checkpoint regulator p53 as well as its negative regulator, the oncogenic mouse double minute-2 homolog protein (MDM2) (Gotoh et al., 2014, 2015). Multiple layers of regulation connect MDM2 function with p53 stability and subcellular localization under normal conditions and in response to genotoxic stress. Briefly, binding of MDM2 to the N-terminal transactivation domain in p53 promotes either monoubiquitylation and nuclear export of p53 or polyubiquitylation and degradation by the 26S proteasome when p53 localizes in the cytosol (Kruse and Gu, 2009). As is the case with other post-translational modifications, ubiquitylation is a reversible process that, in the case of p53, is mediated by the deubiquitylating enzyme herpes virus-associated ubiquitin-specific protease (HAUSP, for review see Kruse and Gu, 2009).

Identification of p53 as a direct interactor of PER2 was first reported using a two-hybrid bacterial screening, confirmed by immunoprecipitation of the endogenous PER2:p53 complex, and proven by *in vitro* competition assays in various cell lines (Gotoh et al., 2014). Binding of PER2 occurs in p53's C-terminus domain, which encompasses nuclear localization signals, both export and import, and the tetramerization domain required for p53 oligomerization and transcriptional activity (Gotoh et al., 2014). Interestingly, binding of PER2 to p53 does not preclude MDM2 recruitment to p53's N-terminus domain and the existence of an endogenous MDM2:p53:PER2 trimeric complex was confirmed (Gotoh et al., 2014). From a functional standpoint, formation of the trimeric complex prevents MDM2-mediated ubiquitylation of p53 (Gotoh et al., 2014). This observation resulted from studies in which p53 ubiquitylation reactions were reconstituted in the presence of PER2 and from purified components *in vitro* (Gotoh et al., 2014). In others, p53's half-life was estimated from cells where the endogenous level of PER2 was either up- or down-regulated (Gotoh et al., 2014). As a result, an initial model suggested that PER2 binding to p53 ensures that basal levels of the tumor suppressor exist for an acute response to, for example, genotoxic stress.

Subcellular distribution of PER2, p53, and MDM2 showed that whereas PER2:p53 localizes both in the cytosol and the nuclear compartments, MDM2:p53:PER2 complex remained solely in the nucleus (Gotoh et al., 2015). Gene expression studies have shown that dissociation of PER2 from p53 in the nucleus is an obligated step for p53 transcriptional activation (Gotoh et al., 2014, 2015). Accordingly, a chimera protein in which PER2 and p53 were covalently linked and folded was unable to modulate the expression of p53 downstream target genes (e.g., *SFN*, *BAX*, *CDKN1a*, *GADD45a*) (Gotoh et al., 2015). Furthermore, the effect of the chimera was achieved even in response to radiation and without compromising the upstream activation of the checkpoint response, i.e., the phosphorylation of ATM/ATR and CHK1/2 (Gotoh et al., 2015). In a related topic of yet unknown biological significance, overexpression of PER2 was shown to promote p53 transcription (Gotoh et al., 2014). Interestingly, the converse regulation in which p53 transcriptionally controls *PER2* gene expression has also been reported and linked to circadian behavior in animals (Miki et al., 2013).

An unexpected observation resulted from the analysis of PER2 and p53 oscillatory behavior in total cell extracts (Gotoh et al., 2016). Based on the finding that PER2 binds and stabilizes p53 (**Figure 1A**; Gotoh et al., 2014, 2015), the conventional wisdom would have been that PER2 and p53 levels oscillate in phase (**Figure 1B**). Instead, PER2 and p53 were found largely out-of-phase relative to each other in cytosolic fractions and matching cell lysates but in-phase in the nucleus (Gotoh et al., 2016). These, *a priori*, paradoxical findings prompted the development of a unique combined theoretical and experimental approach to shed light on the potential scenarios that could reconcile the current experimental data.

An Unbiased Systematic Modeling Approach

We first asked what type of regulation, if any, does PER2 exert over p53 that results in an out-of-phase relationship between these proteins? Which, among the many regulatory processes in the cell that impact p53's production, destruction, localization, and function are modulated by PER2's biology? The relevance of these questions coincides with the complexity of their potential

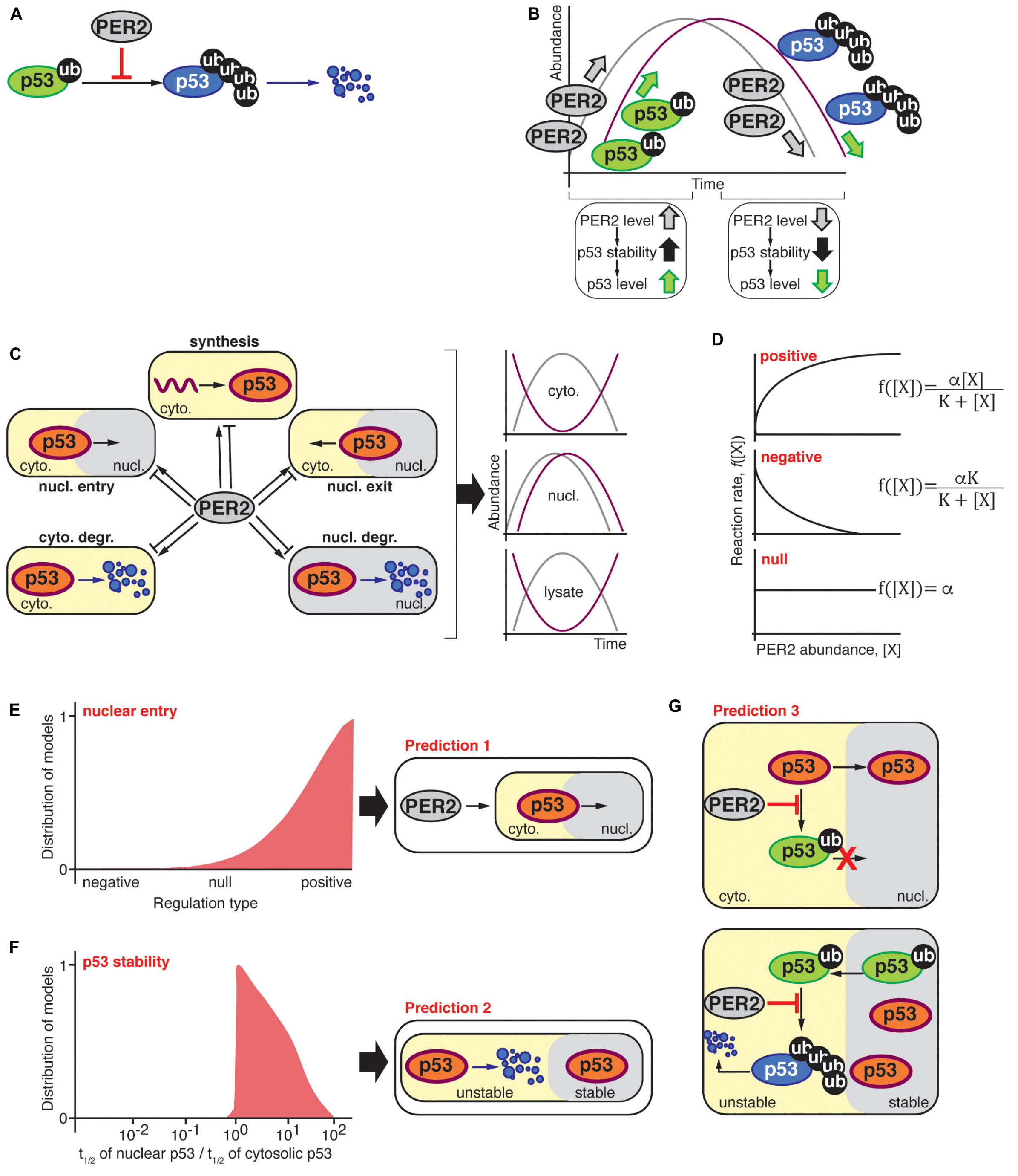


FIGURE 1 | (A) PER2 inhibits MDM2-mediated polyubiquitylation of p53 resulting in p53 stabilization (Gotoh et al., 2014); thus, the abundance of PER2 and p53 were expected to oscillate in-phase **(B)**. **(C)** Schematic representation of possible scenarios by which PER2 could influence p53's phase, i.e., p53 synthesis, shuttling, and degradation in either the cytosol or nuclear compartments. The oscillatory behavior of PER2 and p53 was significantly out-of-phase in both lysate and cytosolic fractions; but in-phase in the nucleus. **(D)** PER2-mediated regulation types for p53 synthesis, shuttling, and degradation were randomly selected among positive, negative, or null in agreement with experimentally determined PER2 and p53 phase relationships during the process of fitting (Gotoh et al., 2016). **(E,F)** Distribution of 10^3 models that successfully simulated the phase relationship between p53 and PER2. The distribution of models is positively skewed toward a PER2-mediated regulation of p53 nuclear entry **(E)**. In addition, simulated fractions of p53's half-life in nuclear and cytosolic compartments resulted in a value larger than one and, thus, p53 was predicted to be more stable in the nuclear compartment **(F)**. Predictions **(E,F)** were incorporated into a comprehensive mathematical model that considered potential molecular mechanisms for PER2 and p53 interaction. Specifically, PER2 promotes p53 nuclear entry by preventing p53's mono-ubiquitylation **(E)**. Once in the nucleus, p53 is mono-ubiquitinated and exits to the cytoplasm, where it is poly-ubiquitinated and degraded **(F)**. As a result, the comprehensive model predicted that PER2 binding to p53 would occur regardless of p53's ubiquitylation status to simulate the phase relationship between p53 and PER2 **(G)**.

answers as a number of combinatorial scenarios that arise from considering all possible interactions and regulations surpass the capabilities of mathematical modeling and high-performance computing facilities. Furthermore, finding an explanation to the out-of-phase paradox was seriously underdetermined by the limited data available -only a handful of oscillatory timeseries for PER2 and p53.

A breakthrough came out when we simplified the p53's complex regulatory network. Accordingly, the hundreds of biochemical processes that modulate p53 activity were categorized under common themes including p53's synthesis, nuclear entry, nuclear exit, cytosolic degradation, and nuclear degradation (**Figure 1C**). Thus, the earlier questions were reduced to a simplified problem that could be tackled using ordinary differential equations and *in silico* experiments. To simulate all possible scenarios by which PER2 could influence production, shuttling, or degradation of p53 that resulted in in-phase distribution of PER and p53 in the nucleus but anti-phase in the cytosol, we derived the following equation:

$$\begin{aligned}\frac{d[p53_c]}{dt} &= f_1([Per_n]) - f_2([Per_c])[p53_c] + f_4([Per_n])[p53_n] \\ &\quad - f_3([Per_c])\frac{[p53_c]}{K_{\text{deg}} + [p53_c]} \\ \frac{d[p53_n]}{dt} &= f_2([Per_c])[p53_c] - f_4([Per_n])[p53_n] \\ &\quad - f_5([Per_n])\frac{[p53_n]}{K_{\text{deg}} + [p53_n]} \end{aligned} \quad (1)$$

where $[p53_c]$ and $[Per_c]$, $[p53_n]$ and $[Per_n]$ represent the concentration of p53 and PER2 in the cytoplasm and nucleus, respectively. In addition, oscillations in $[Per_c]$ and $[Per_n]$ were simulated *via* the Kim and Forger model, a mathematical model of the intracellular mammalian circadian clock (Kim and Forger, 2012; Kim et al., 2014; Kim, 2016).

Unlike most mathematical models, the functions f_1, f_2, f_3, f_4, and f_5 were not specified in Equation (1). Each function describes a type of regulation on the production, nuclear import, cytosolic degradation, nuclear export, and nuclear degradation of p53, respectively, that is mediated by PER2 (**Figure 1D**). Regulation types were randomly selected among three function f_i forms:

$$f_i([X]) = \frac{\alpha_i[X]}{K_i + [X]}, \quad f_i([X]) = \frac{\alpha_i K_i}{K_i + [X]}, \quad \text{or} \quad f_i([X]) = \alpha_i \quad (2)$$

which represent positive, negative, or null regulations *via* PER2 [$[X]$ in Equation (2)], respectively. This approach allowed for the investigation of all possible scenarios for which PER2 mediates p53 regulation.

After unbiased selection of f_i functions (i.e., one-third for each regulation type of positive, null, or negative interactions), the value of the unknown parameters for these functions, i.e., K_{deg} in Equation (1) and α_i, and K_i in Equation (2), were estimated using a global stochastic parameter search algorithm to simulate the phase relationship between PER2 and p53 (see Gotoh et al., 2016 for details). Of the $\sim$1 million simulated scenarios, 1,000 models successfully predicted the correct PER2 and p53 phase relationships in the cytosol and nucleus. Even when the selection of the regulation type was unbiased, successful cases showed a strong skewness toward a positive regulation of nuclear entry (**Figure 1E**). This implies that PER2-mediated p53 nuclear entry is essential to accurately simulate the proteins' phase relationship (Prediction 1). Further analysis of the unknown parameter values (α_i, and K_i and K_{deg}) unveiled another unifying property among successful cases: p53 should be more stable in the nucleus than in the cytoplasm (Prediction 2, **Figure 1F**).

Predictions 1 and 2 arise from simplifying the biological system and neglecting detailed molecular mechanisms. Consequently, the model informs us about the "events," e.g., PER2 shuttles p53 to the nucleus or p53 stability is altered in different compartments, but not about how the events are driven. For example, how does PER2 promote p53 nuclear entry? Or how is p53 stability achieved? Answers to the "hows" were arrived at by incorporating two key pieces of information from previous studies: (*i*) PER2 binding to p53 inhibits p53's ubiquitylation (**Figure 1A**; Gotoh et al., 2014), and (*ii*) p53 ubiquitylation status modulates its shuttling and degradation (O'Brate and Giannakakou, 2003; Gotoh et al., 2016). By integrating both the latest information and the model's predictions, a more detailed and realistic mathematical model emerged. As a result, a newly refined model included the PER transcriptional negative feedback loop and considered the various ubiquitylated forms of p53 along with the interaction between PER2 and p53; all of which, were summarized in 13 differential equations containing 18 parameters that were thoroughly described in Gotoh et al. (2016).

Two assumptions were at the foundation of the newly refined model: (*i*) PER2 promotes p53's nuclear entry by blocking p53's ubiquitylation within the nuclear import motif (Marchenko et al., 2010) and, (*ii*) once in the cytoplasm, mono-ubiquitylated p53 becomes poly-ubiquitylated and thus, unstable (**Figure 1G**; Marchenko et al., 2010). Along with restrictions in the parameters based on the aforementioned assumptions, parameters were estimated by the simulated annealing method. This allowed for the refined model to simulate the phase relationship between the two proteins. Following the analysis of the estimated parameters, we found that PER2:p53 binding should occur, regardless of p53's ubiquitylation status, to properly simulate the antiphase relationship between PER2 and p53 that was reported to exist in total cell lysates (Prediction 3).

From *in silico* Prediction to *in vitro* Validation

The three *in silico* predictions were next challenged in a cell-based system (Gotoh et al., 2016). Accordingly, experiments were devoted to test the prediction that PER2 binding to p53 promotes p53's nuclear translocation (**Figure 1E**; prediction 1). The rationale for an experimental design was conceptually simple: If PER2 were to favor p53 translocation, then PER2 ectopic expression, or its downregulation, should impact the accumulation of p53 in either cellular compartment. In accordance with this premise, PER-mediated nuclear shuttling of p53 was confirmed in cells overexpressing trace levels of PER2 and maintained in the presence of both proteasome and nuclear export inhibitors (Gotoh et al., 2016). The converse experiment, in which endogenous PER2 was downregulated, also resulted in a reduction in the total level of p53 present in the nucleus but not its complete abrogation (Gotoh et al., 2016), suggesting that PER2 certainly contributes to p53 shuttling but that other routes exist.

Other work evaluated the half-life of p53 in purified nuclear and cytosolic extracts obtained from cells treated with cycloheximide, a protein translation inhibitor. In agreement with the theoretical model (**Figure 1F**; prediction 2), the half-life of endogenous p53 was ∼sevenfold longer when localized in the nucleus than in the cytosolic compartment and, thus, p53's localization influences the kinetics of its degradation (Gotoh et al., 2016). Interestingly, PER2's half-life remained comparable in both compartments.

Gotoh et al. (2014) established that PER2 binding to p53 prevents p53's polyubiquitylation favoring its stability. This finding posed a provocative prediction that PER2 should be, nevertheless, able to bind a form of p53 containing multiple ubiquitin moieties (**Figure 1G**; prediction 3). Stepwise *in vitro* assays using recombinant purified proteins showed that, indeed, PER2 binds to mono and polyubiquitylated p53 (although, the extent to which the various forms of PER2:p53 complexes interplay remains to be explored) (Gotoh et al., 2016). As a result, the systematic mathematical approach helped to reconcile seemingly contradictory experimental data by generating theoretical predictions that were, then, experimentally confirmed.

A Bidirectional Relationship Between Clock and Checkpoint Components

The interaction of PER2 with p53 and MDM2 brings to our attention the need for understanding the role that circadian proteins play at critical nodes of regulation in the cell, e.g., at the checkpoint crosstalk. Conversely, it poses the question of whether cellular p53 and/or MDM2 components influence the clock itself.

In eukaryotic organisms, the stability of PER2, which is responsive to environmental signals and homeostatic cellular conditions, is driven by multiple phosphorylation events that influence the period length and phase of the circadian rhythm (Gallego and Virshup, 2007; Chiu et al., 2011; Lee et al., 2011; Reischl and Kramer, 2011; D'Alessandro et al., 2015; Zhou et al., 2015). PER2 is targeted for post-translational modification by CK1ε/δ, glycogen synthase kinase 3β (GSK3β), casein kinase 2 (CK2), and casein kinase 1α (CK1α), all of which lead to either PER2 translocation or altered stability (for review see Ode and Ueda, 2018). Support for the role of CK1ε/δ in period determination emerged from multiple sources including genetic/phenotypic screenings in organisms as diverse as flies and mice, high-throughput studies using either small molecule inhibitors or short interfering RNAs, and disease genotyping in the human population (Kloss et al., 1998; Price et al., 1998; Lowrey et al., 2000; Xu et al., 2005; Badura et al., 2007; Hirota et al., 2008; Zhang et al., 2009; Meng et al., 2010; Chen et al., 2012; Kim et al., 2013, 2019). In the case of CK2, the kinase seems to play a dual role by phosphorylating the N-terminus domain of PER2 and by cooperating with CK1ε to promote PER2 degradation in the mammalian clock (Tsuchiya et al., 2009). Unlike CK1, the activity of GSK3β influences period length by targeting PER2 and promoting PER2's nuclear translocation (Iitaka et al., 2005). Thus, PER2's stability, shuttling, and the activity of various kinases are all part of an entangled network involved in period determination.

Key molecular events define the best understood mechanism for PER2 turnover; these include, CK1ε/δ substrate phosphorylation followed by the E3 ligase β-transducin repeat-containing protein 1/2 (β-TrCP1/2) recognition of PER2, ubiquitylation, and PER2's proteasomal degradation (Eide et al., 2005; Ohsaki et al., 2008). However, a number of experimental observations suggested that additional mechanisms might exist. For example, co-expression of dominant negative forms of β-TrCP1 and β-TrCP2 in cells stabilize PER2 rather than promote its degradation (Ohsaki et al., 2008). Results reported by Walton et al. (2009) and Lee et al. (2011). were in the same line in which neither the expression of a dominant negative form of the CK1ε isoform in a CK1$\delta^{-/-}$ cellular background nor the pharmacological inhibition of CK1δ/ε isoforms, respectively, completely abrogated the circadian oscillation of a reporter gene. More recently, Zhou et al. proposed a phosphoswitch model to explain the three-stage kinetic of PER2 degradation that occurs during the PER2's accumulation phase (Zhou et al., 2015; Narasimamurthy et al., 2018). Unlike the rapid initial decay and second accumulation stage, PER2's degradation in both the third stage and falling phase of the circadian oscillation were independent of phosphorylation and β-TrCP activity (Zhou et al., 2015). Finally, ubiquitylated PER proteins were also detected in β-TrCP1/2 knockdown cells, although PER degradation occurred at a slower rate than in wild type cells (D'Alessandro et al., 2017). The above cumulative evidence suggests that a phosphorylation-independent and ubiquityl-mediated turnover of PER2 could exist. Thus, a role for MDM2 in modulating PER2 stability is plausible.

Identification of PER2 as a previously uncharacterized substrate of MDM2 was shown to occur in the absence of p53 binding (Liu et al., 2018). First, PER2:MDM2 association was detected by protein complementation and later confirmed by immunoprecipitation of the endogenous complex in various cellular systems (Liu et al., 2018). Conformational epitope mapping showed that PER2 binding occurs downstream of the p53-binding domain and upstream of the RING domain in MDM2 (Liu et al., 2018). This finding is of relevance when considering the spatial assembly of the PER2:p53:MDM2 complex in a context in which structural information of either component alone or in association with each other is missing. By binding downstream of the p53-binding domain, PER2 association to MDM2 makes it possible for p53 to bind to, and be part of, the PER2:p53:MDM2 complex as it was initially shown to exist in the nucleus (Gotoh et al., 2015). When PER2 bound to the N-terminus of the RING-domain, MDM2 was still able to exert its RING-dependent ubiquitin ligase activity toward its substrates. Accordingly, MDM2-mediated ubiquitylation of PER2 was shown to occur and be preferentially mediated by UbcH5a at the ubiquitin linkages Lys^{11} and Lys^{48} (Liu et al., 2018).

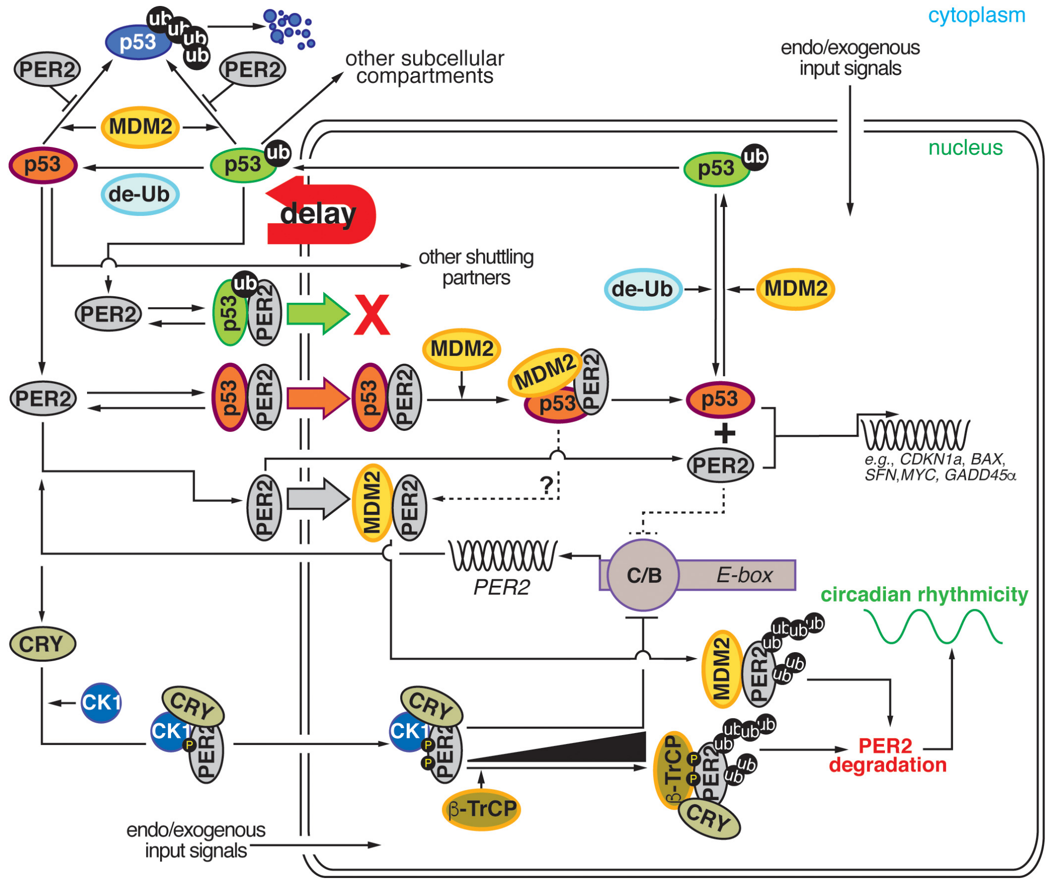


FIGURE 2 | Comprehensive diagram representing the existing bidirectional interaction within the PER2:p53:MDM2 network. Cartoon representation depicting the four major regulatory nodes in p53, PER2, and MDM2 signaling. Node 1: PER2 modulates p53 stability by preventing MDM2-mediated ubiquitylation and proteasomal degradation of p53. Node 2: PER2:p53 binding favors p53 shuttling to the nucleus where it binds MDM2 and forms a stable trimeric complex. Node 3: Dissociation of PER2:p53 facilitates each component's transcriptional activity and expression of downstream target genes. Node 4: Accumulation of PER2:MDM2 in the nucleus results in PER2 polyubiquitylation and degradation in a process that influences circadian period length. de-Ub, de-ubiquitylase; ub, ubiquitin; CK1δ/ε, casein kinase 1 δ/ε; CRY, cryptochrome; β-TrCP, β-transducin repeat-containing protein; PER2, period 2; MDM2, mouse double minute-2 homolog protein; C/B, circadian locomotor output cycles kaput/BMAL1 (Brain and Muscle ARNT-like (CLOCK/BMAL1); E-box, Enhancer element; g-IR, irradiation. Delay: represents the phase delay that results from PER2-mediated p53 shuttling.

Interestingly, polyubiquitylation sites in PER2 mapped within the interface of binding between PER2 and p53, a result that suggests PER2:p53 association might serve the purpose of stabilizing both components of the complex (Liu et al., 2018). Unlike the case of β-TrCP, where substrate phosphorylation was established to be an absolute requirement for β-TrCP recognition and ligase activity, phosphorylation of PER2 was neither a pre-requisite for MDM2 binding nor for MDM2-mediated ubiquitylation as shown in cultured cells and in a recombinant system (Liu et al., 2018). Corresponding with the role of polyubiquitylation signals in proteasomal degradation, the activity of MDM2 toward PER2 shortened PER2's half-life whereas MDM2 down-regulation or its chemical inhibition favors PER2 stability (Liu et al., 2018). Accordingly, ectopic expression of MDM2 shortened the period length where its downregulation, or inhibition, resulted in a significant lengthening of the circadian period, which suggests that control over MDM2 activity is a relevant node of circadian regulation (Liu et al., 2018).

The interplay between p53, MDM2, and PER2 establishes a node of regulation where circadian and checkpoint components can bidirectionally communicate and influence each other's downstream signaling. As depicted in **Figure 2**, network connections can be clustered around five emerging themes: (*i*) PER2's control of p53 stability, (*ii*) PER2's regulation of nuclear p53 shuttling, (*iii*) PER2's tuning of p53's transcriptional activity, (*iv*) PER2's interaction with MDM2, and (*v*) MDM2's targeting of PER2 for degradation. All of these themes should be evaluated in the context of the well-established regulatory relationship between p53 and MDM2 (for review see Kruse and Gu, 2009; Vousden and Prives, 2009). In this relationship (*i*) p53 and MDM2 interact with each other, (*ii*) p53 protein levels oscillate in the cell, (*iii*) MDM2 mono-ubiquitylates p53 in the nucleus, (*iv*) the mono-ubiquitylated complex translocates to the cytosol where (*v*) p53 is polyubiquitiylated, and (*vi*) degraded by the proteasome. As a result, the picture that emerges is one in which a bidirectional regulation exists and occurs at multiple levels to ensure that events take place at specific times in the day and that enough time is given to the cell to respond to environmental and intracellular perturbations (**Figure 2**).

CONCLUDING REMARKS

This work established that molecular connections between circadian and cell cycle checkpoint components occur bidirectionally. Quality research, as well as in-depth biochemical, molecular, structural, and behavioral studies in animal models aided by innovative modeling strategies, has helped in identifying previously unexpected crosstalk regulatory processes. This leaves us with the opportunity to tackle fundamental questions whose answers would certainly carry weight in the translational arena. For example, as circadian proteins interact with oncogene and tumor suppressor proteins, would the response to genotoxic stress be different in different circadian phases? Should therapeutic approaches aimed at targeting p53 be administered at specific times of the day? Is circadian rhythm disruption in response to DNA damage the consequence of changes in MDM2's activity and, potentially, PER2's stability? Answers to these questions lay in adequately retrieving system-level information and extending the complexity of existing models, an achievable goal with societal benefits in times where interdisciplinary research is being fostered.

AUTHOR CONTRIBUTIONS

XZ, DK, TG, and JL provided feedback and participated in the design of the figures. JK and CF wrote the manuscript and edited the figures. XZ, TG, JL, JK, and CF participated in the original research presented in the manuscript. All authors have read and approved the final version of this manuscript.

FUNDING

This work was supported by the National Science Foundation MCB division (MCB-1517298) to CF, VT Open Access Subvention Fund, and Human Frontiers Science Program Organization (RGY0063/2017), National Research Foundation of Korea Grant (NRF-2016 RICIB 3008468) to JK, and NRF-2017-Fostering Core Leaders of the Future Basic Science Program/Global Ph.D. Fellowship Program to DK.

ACKNOWLEDGMENTS

We apologize to colleagues whose work has not been cited due to space limitations. We thank DGS Capelluto for critical reading of the manuscript and all members of the Finkielstein laboratory for help and discussions. We would also like to thank Dr. J. Webster for comments and manuscript editing.

REFERENCES

Badura, L., Swanson, T., Adamowicz, W., Adams, J., Cianfrogna, J., Fisher, K., et al. (2007). An inhibitor of casein kinase I epsilon induces phase delays in circadian rhythms under free-running and entrained conditions. *J. Pharmacol. Exp. Ther.* 322, 730–738. doi: 10.1124/jpet.107.122846

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., et al. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat. Rev. Genet.* 6, 544–556. doi: 10.1038/nrg1633

Bjarnason, G. A., Jordan, R. C., and Sothern, R. B. (1999). Circadian variation in the expression of cell-cycle proteins in human oral epithelium. *Am. J. Pathol.* 154, 613–622. doi: 10.1016/S0002-9440(10)65306-65300

Bjarnason, G. A., Jordan, R. C., Wood, P. A., Li, Q., Lincoln, D. W., Sothern, R. B., et al. (2001). Circadian expression of clock genes in human oral mucosa and skin: association with specific cell-cycle phases. *Am. J. Pathol.* 158, 1793–1801. doi: 10.1016/S0002-9440(10)64135-64131

Brown, W. R. (1991). A review and mathematical analysis of circadian rhythms in cell proliferation in mouse, rat, and human epidermis. *J. Invest. Dermatol.* 97, 273–280. doi: 10.1111/1523-1747.ep12480379

Chen, Z., Yoo, S. H., Park, Y. S., Kim, K. H., Wei, S., Buhr, E., et al. (2012). Identification of diverse modulators of central and peripheral circadian clocks by high-throughput chemical screening. *Proc. Natl. Acad. Sci. U.S.A.* 109, 101–106. doi: 10.1073/pnas.1118034108

Chiu, J. C., Ko, H. W., and Edery, I. (2011). NEMO/NLK phosphorylates PERIOD to initiate a time-delay phosphorylation circuit that sets circadian clock speed. *Cell* 145, 357–370. doi: 10.1016/j.cell.2011.04.002

D'Alessandro, M., Beesley, S., Kim, J. K., Chen, R., Abich, E., Cheng, W., et al. (2015). A tunable artificial circadian clock in clock-defective mice. *Nat. Commun.* 6:8587. doi: 10.1038/ncomms9587

D'Alessandro, M., Beesley, S., Kim, J. K., Jones, Z., Chen, R., Wi, J., et al. (2017). Stability of wake-sleep cycles requires robust degradation of the PERIOD protein. *Curr. Biol.* 27, 3454–3467.e8. doi: 10.1016/j.cub.2017.10.014

Destici, E., Oklejewicz, M., Saito, S., and van der Horst, G. T. (2011). Mammalian cryptochromes impinge on cell cycle progression in a circadian clock-independent manner. *Cell Cycle* 10, 3788–3797. doi: 10.4161/cc.10.21.17974

Eide, E. J., Woolf, M. F., Kang, H., Woolf, P., Hurst, W., Camacho, F., et al. (2005). Control of mammalian circadian rhythm by CKIepsilon-regulated proteasome-mediated PER2 degradation. *Mol. Cell Biol.* 25, 2795–2807. doi: 10.1128/MCB.25.7.2795-2807.2005

Gallego, M., and Virshup, D. M. (2007). Post-translational modifications regulate the ticking of the circadian clock. *Nat. Rev. Mol. Cell Biol.* 8, 139–148. doi: 10.1038/nrm2106

Gaucher, J., Montellier, E., and Sassone-Corsi, P. (2018). Molecular cogs: interplay between circadian clock and cell cycle. *Trends Cell Biol.* 28, 368–379. doi: 10.1016/j.tcb.2018.01.006

Gery, S., Komatsu, N., Baldjyan, L., Yu, A., Koo, D., and Koeffler, H. P. (2006). The circadian gene per1 plays an important role in cell growth and DNA damage control in human cancer cells. *Mol. Cell* 22, 375–382. doi: 10.1016/j.molcel.2006.03.038

Gotoh, T., Kim, J. K., Liu, J., Vila-Caballer, M., Stauffer, P. E., Tyson, J. J., et al. (2016). Model-driven experimental approach reveals the complex regulatory distribution of p53 by the circadian factor Period 2. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13516–13521. doi: 10.1073/pnas.1607984113

Gotoh, T., Vila-Caballer, M., Liu, J., Schiffhauer, S., and Finkielstein, C. V. (2015). Association of the circadian factor Period 2 to p53 influences p53's function in DNA-damage signaling. *Mol. Biol. Cell* 26, 359–372. doi: 10.1091/mbc.E14-05-0994

Gotoh, T., Vila-Caballer, M., Santos, C. S., Liu, J., Yang, J., and Finkielstein, C. V. (2014). The circadian factor Period 2 modulates p53 stability and transcriptional activity in unstressed cells. *Mol. Biol. Cell* 25, 3081–3093. doi: 10.1091/mbc.E14-05-0993

Grechez-Cassiau, A., Rayet, B., Guillaumond, F., Teboul, M., and Delaunay, F. (2008). The circadian clock component BMAL1 is a critical regulator of p21WAF1/CIP1 expression and hepatocyte proliferation. *J. Biol. Chem.* 283, 4535–4542. doi: 10.1074/jbc.M705576200

Hirota, T., Lewis, W. G., Liu, A. C., Lee, J. W., Schultz, P. G., and Kay, S. A. (2008). A chemical biology approach reveals period shortening of the mammalian circadian clock by specific inhibition of GSK-3beta. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20746–20751. doi: 10.1073/pnas.0811410106

Hong, C. I., Zamborszky, J., Baek, M., Labiscsak, L., Ju, K., Lee, H., et al. (2014). Circadian rhythms synchronize mitosis in *Neurospora crassa*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1397–1402. doi: 10.1073/pnas.1319399111

Huber, A. L., Papp, S. J., Chan, A. B., Henriksson, E., Jordan, S. D., Kriebs, A., et al. (2016). CRY2 and FBXL3 cooperatively degrade c-MYC. *Mol. Cell* 64, 774–789. doi: 10.1016/j.molcel.2016.10.012

Iitaka, C., Miyazaki, K., Akaike, T., and Ishida, N. (2005). A role for glycogen synthase kinase-3beta in the mammalian circadian clock. *J. Biol. Chem.* 280, 29397–29402. doi: 10.1074/jbc.M503526200

Kim, D. W., Chang, C., Chen, X., Doran, A. C., Gaudreault, F., Wager, T., et al. (2019). Systems approach reveals photosensitivity and PER2 level as determinants of clock-modulator efficacy. *Mol. Syst. Biol.* 15:e8838. doi: 10.15252/msb.20198838

Kim, J. K. (2016). Protein sequestration versus Hill-type repression in circadian clock models. *IET Syst. Biol.* 10, 125–135. doi: 10.1049/iet-syb.2015.0090

Kim, J. K., and Forger, D. B. (2012). A mechanism for robust circadian timekeeping via stoichiometric balance. *Mol. Syst. Biol.* 8:630. doi: 10.1038/msb.2012.62

Kim, J. K., Forger, D. B., Marconi, M., Wood, D., Doran, A., Wager, T., et al. (2013). Modeling and validating chronic pharmacological manipulation of circadian rhythms. *CPT Pharmacomet. Syst. Pharmacol.* 2:e57. doi: 10.1038/psp.2013.34

Kim, J. K., Josic, K., and Bennett, M. R. (2014). The validity of quasi-steady-state approximations in discrete stochastic simulations. *Biophys. J.* 107, 783–793. doi: 10.1016/j.bpj.2014.06.012

Kloss, B., Price, J. L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C. S., et al. (1998). The Drosophila clock gene double-time encodes a protein closely related to human casein kinase Iepsilon. *Cell* 94, 97–107. doi: 10.1016/s0092-8674(00)81225-81228

Koike, N., Yoo, S. H., Huang, H. C., Kumar, V., Lee, C., Kim, T. K., et al. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338, 349–354. doi: 10.1126/science.1226339

Kowalska, E., Ripperger, J. A., Hoegger, D. C., Bruegger, P., Buch, T., Birchler, T., et al. (2013). NONO couples the circadian clock to the cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1592–1599. doi: 10.1073/pnas.1213317110

Kruse, J. P., and Gu, W. (2009). Modes of p53 regulation. *Cell* 137, 609–622. doi: 10.1016/j.cell.2009.04.050

Lee, H. M., Chen, R., Kim, H., Etchegaray, J. P., Weaver, D. R., and Lee, C. (2011). The period of the circadian oscillator is primarily determined by the balance between casein kinase 1 and protein phosphatase 1. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16451–16456. doi: 10.1073/pnas.1107178108

Liu, J., Zou, X., Gotoh, T., Brown, A. M., Jiang, L., Wisdom, E. L., et al. (2018). Distinct control of PERIOD2 degradation and circadian rhythms by the oncoprotein and ubiquitin ligase MDM2. *Sci. Signal.* 11:556. doi: 10.1126/scisignal.aau0715

Lowrey, P. L., Shimomura, K., Antoch, M. P., Yamazaki, S., Zemenides, P. D., Ralph, M. R., et al. (2000). Positional syntenic c. *Science* l, 483–492.

Marchenko, N. D., Hanel, W., Li, D., Becker, K., Reich, N., and Moll, U. M. (2010). Stress-mediated nuclear stabilization of p53 is regulated by ubiquitination and importin-alpha3 binding. *Cell Death Differ.* 17, 255–267. doi: 10.1038/cdd.2009.173

Matsuo, T., Yamaguchi, S., Mitsui, S., Emi, A., Shimoda, F., and Okamura, H. (2003). Control mechanism of the circadian clock for timing of cell division in vivo. *Science* 302, 255–259. doi: 10.1126/science.1086271

Matsu-Ura, T., Dovzhenok, A., Aihara, E., Rood, J., Le, H., Ren, Y., et al. (2016). Intercellular coupling of the cell cycle and circadian clock in adult stem cell culture. *Mol. Cell* 64, 900–912. doi: 10.1016/j.molcel.2016.10.015

Mazzoccoli, G., Laukkanen, M. O., Vinciguerra, M., Colangelo, T., and Colantuoni, V. (2016). A timeless link between circadian patterns and disease. *Trends Mol. Med.* 22, 68–81. doi: 10.1016/j.molmed.2015.11.007

Meng, Q. J., Maywood, E. S., Bechtold, D. A., Lu, W. Q., Li, J., Gibbs, J. E., et al. (2010). Entrainment of disrupted circadian behavior through inhibition of casein kinase 1 (CK1) enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15240–15245. doi: 10.1073/pnas.1005101107

Miki, T., Matsumoto, T., Zhao, Z., and Lee, C. C. (2013). p53 regulates Period2 expression and the circadian clock. *Nat. Commun.* 4:2444. doi: 10.1038/ncomms3444

Mori, T., Binder, B., and Johnson, C. H. (1996). Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10183–10188. doi: 10.1073/pnas.93.19.10183

Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F., and Schibler, U. (2004). Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* 119, 693–705. doi: 10.1016/j.cell.2004.11.015

Narasimamurthy, R., Hunt, S. R., Lu, Y., Fustin, J. M., Okamura, H., Partch, C. L., et al. (2018). CK1delta/epsilon protein kinase primes the PER2 circadian phosphoswitch. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5986–5991. doi: 10.1073/pnas.1721076115

O'Brate, A., and Giannakakou, P. (2003). The importance of p53 location: nuclear or cytoplasmic zip code? *Drug Resis.t Updat.* 6, 313–322.

Ode, K. L., and Ueda, H. R. (2018). Design principles of phosphorylation-dependent timekeeping in eukaryotic circadian clocks. *Cold Spring Harb. Perspect. Biol.* 10:a028357. doi: 10.1101/cshperspect.a028357

Ohsaki, K., Oishi, K., Kozono, Y., Nakayama, K., Nakayama, K. I., and Ishida, N. (2008). The role of {beta}-TrCP1 and {beta}-TrCP2 in circadian rhythm generation by mediating degradation of clock protein PER2. *J. Biochem.* 144, 609–618. doi: 10.1093/jb/mvn112

Papp, S. J., Huber, A. L., Jordan, S. D., Kriebs, A., Nguyen, M., Moresco, J. J., et al. (2015). DNA damage shifts circadian clock time via Hausp-dependent Cry1 stabilization. *eLife* 4:e04883. doi: 10.7554/eLife.04883

Price, J. L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., and Young, M. W. (1998). double-time is a novel Drosophila clock gene that regulates PERIOD protein accumulation. *Cell* 94, 83–95. doi: 10.1016/s0092-8674(00)81224-81226

Reischl, S., and Kramer, A. (2011). Kinases and phosphatases in the mammalian circadian clock. *FEBS Lett.* 585, 1393–1399. doi: 10.1016/j.febslet.2011.02.038

Takahashi, J. S. (2017). Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* 18, 164–179. doi: 10.1038/nrg.2016.150

Tamai, T. K., Young, L. C., Cox, C. A., and Whitmore, D. (2012). Light acts on the zebrafish circadian clock to suppress rhythmic mitosis and cell proliferation. *J. Biol. Rhythms* 27, 226–236. doi: 10.1177/0748730412440861

Tsuchiya, Y., Akashi, M., Matsuda, M., Goto, K., Miyata, Y., Node, K., et al. (2009). Involvement of the protein kinase CK2 in the regulation of mammalian circadian rhythms. *Sci. Signal.* 2:ra26. doi: 10.1126/scisignal.2000305

Unsal-Kacmaz, K., Mullen, T. E., Kaufmann, W. K., and Sancar, A. (2005). Coupling of human circadian and cell cycles by the timeless protein. *Mol. Cell. Biol.* 25, 3109–3116. doi: 10.1128/MCB.25.8.3109-3116.2005

Vousden, K. H., and Prives, C. (2009). Blinded by the light: the growing complexity of p53. *Cell* 137, 413–431. doi: 10.1016/j.cell.2009.04.037

Walton, K. M., Fisher, K., Rubitski, D., Marconi, M., Meng, Q. J., Sladek, M., et al. (2009). Selective inhibition of casein kinase 1 epsilon minimally alters circadian clock period. *J. Pharmacol. Exp. Ther.* 330, 430–439. doi: 10.1124/jpet.109.151415

Xu, Y., Padiath, Q. S., Shapiro, R. E., Jones, C. R., Wu, S. C., Saigoh, N., et al. (2005). Functional consequences of a CKIdelta mutation causing familial advanced sleep phase syndrome. *Nature* 434, 640–644. doi: 10.1038/nature03453

Yang, Q., Pando, B. F., Dong, G., Golden, S. S., and van Oudenaarden, A. (2010). Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327, 1522–1526. doi: 10.1126/science.1181759

Zhang, E. E., Liu, A. C., Hirota, T., Miraglia, L. J., Welch, G., Pongsawakul, P. Y., et al. (2009). A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* 139, 199–210. doi: 10.1016/j.cell.2009.08.031

Zhou, M., Kim, J. K., Eng, G. W., Forger, D. B., and Virshup, D. M. (2015). A Period2 phosphoswitch regulates and temperature compensates circadian period. *Mol. Cell* 60, 77–88. doi: 10.1016/j.molcel.2015.08.022

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 10 July 2020
doi: 10.3389/fphys.2020.00822

Stochastic Methods for Inferring States of Cell Migration

R. J. Allen[1]†, C. Welch[1]†, Neha Pankow[1], Klaus M. Hahn[1,2]* and Timothy C. Elston[1,2]*

[1] Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2] Computational Medicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
David McMillen,
University of Toronto Mississauga, Canada
Nathan Weinstein,
Universidad Nacional Autónoma de México, Mexico

***Correspondence:**
Klaus M. Hahn
khahn@med.unc.edu
Timothy C. Elston
telston@med.unc.edu

†Present address:
R. J. Allen,
Early Clinical Development, Quantitative Systems Pharmacology, Pfizer Inc., Cambridge, MA, United States
C. Welch,
Otolarynology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, NC, United States

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: 04 April 2020
Accepted: 19 June 2020
Published: 10 July 2020

Citation:
Allen RJ, Welch C, Pankow N, Hahn KM and Elston TC (2020) Stochastic Methods for Inferring States of Cell Migration. Front. Physiol. 11:822. doi: 10.3389/fphys.2020.00822

Cell migration refers to the ability of cells to translocate across a substrate or through a matrix. To achieve net movement requires spatiotemporal regulation of the actin cytoskeleton. Computational approaches are necessary to identify and quantify the regulatory mechanisms that generate directed cell movement. To address this need, we developed computational tools, based on stochastic modeling, to analyze time series data for the position of randomly migrating cells. Our approach allows parameters that characterize cell movement to be efficiently estimated from cell track data. We applied our methods to analyze the random migration of Mouse Embryonic Fibroblasts (MEFS) and HeLa cells. Our analysis revealed that MEFs exist in two distinct states of migration characterized by differences in cell speed and persistence, whereas HeLa cells only exhibit a single state. Further analysis revealed that the Rho-family GTPase RhoG plays a role in determining the properties of the two migratory states of MEFs. An important feature of our computational approach is that it provides a method for predicting the current migration state of an individual cell from time series data. Finally, we applied our computational methods to HeLa cells expressing a Rac1 biosensor. The Rac1 biosensor is known to perturb movement when expressed at overly high concentrations; at these expression levels the HeLa cells showed two migratory states, which correlated with differences in the spatial distribution of active Rac1.

Keywords: cell migration, stochastic modeling, RHOG, Rac1, biosenor, migration states

INTRODUCTION

The ability of cells to move is essential to many biological processes, such as tissue development, the immune response and wound healing (Franca-Koh et al., 2007; Petrie et al., 2009; Cain and Ridley, 2012). Anomalous cell migration plays a role in diseases, such as cancer and atherosclerosis (Cain and Ridley, 2012; Hall, 2009; Lemarié et al., 2010; Finney et al., 2017). During cell migration, intracellular signaling networks tightly control the spatiotemporal dynamics of the cytoskeleton. In particular, the Rho family of small GTPases has been implicated in membrane protrusion, adhesion, contraction and de-adhesion, all steps necessary for cell migration (Rottner et al., 1999; Jaffe and Hall, 2005; Goley and Welch, 2006; Ridley, 2006; Iden and Collard, 2008; Ladwein and Rottner, 2008). Rac1, the family member studied here, produces cell protrusions by interacting with effector proteins that modulate actin polymerization, including formins and Paks. A prevailing hypothesis is that Rac1 induces localized actin polymerization to trap random, thermal driven outward movements of the cell edge (Ridley, 2015; Marston et al., 2019; Schaks et al., 2019).

During random cell migration, in which cells do not experience directional environmental cues, cells move in a persistent manner, but with significant variability in their direction and speed. Therefore, methods for quantifying cell movement that take into account the stochastic nature of this phenomenon are needed. Previous studies have analyzed cell migration in terms of quantitative metrics such as the mean squared deviation in cell position, which can be linked to both speed and persistence (Othmer et al., 1988; Dimilla et al., 1992; Rosello et al., 2004; Dieterich et al., 2008). Additionally, it has been suggested that fractional diffusion models are required to accurately describe cell movement (Dieterich et al., 2008). We refer the reader to a recent review which describes these approaches and others (Svensson et al., 2018). We used stochastic modeling to develop tools for quantifying cell migration such that it can be characterized in terms of biologically relevant parameters. In our approach, the motion of cells is assumed to follow a 2D random walk with persistence. A related method that takes into account the probability of turning and contains a parameter related to persistence also has been applied to analyze random cell migration (Arrieumerlou and Meyer, 2005). An important distinction of our approach is that our model allows for the possibility of multiple states of migration, distinguished by differences in speed and persistence. This feature allowed us to determine that Mouse Embryonic Fibroblasts (MEFS) exist in two distinct states during random migration. Knock down of the Rho-GTPase RhoG suggests that this protein plays an important role in establishing the two states. We next demonstrated how our method allows the migration state of a cell to be predicted from time series data. Finally, we applied our method to examine the activation of Rac1, a GTPase known to be important in producing localized protrusions. Interestingly, we found that overexpressed, biosensor induced two states of migration in HeLa cells that correlated with different numbers of active Rac1 foci.

RESULTS

Preliminary Analysis

To develop our methods, we collected data sets that consisted of time series for the x and y coordinates of the cell centroids of randomly migrating MEF cells (**Figures 1A,B**). We chose this cell type because it shows persistent migration in the absence of directional cues. As an initial analysis of the data, we computed the average persistence of cell movement defined as $P = <\cos(\theta)>$, where θ is the change in the direction of cell movement between measurements (**Figure 1C**) and the angular bracket denotes averaging over cell tracks. If θ is uniformly distributed, then the motion of the cell lacks persistence and $P = 0$. This behavior would be consistent with a pure random walk (diffusive motion). For values of P greater than zero, the movement of the cell shows persistence, with a value of 1 indicating motion in a straight line. Combining the cell tracks for individual cells, produced a value of $P = 0.43$. This value is consistent with cells that show persistent motion. We also generated histograms from the Δx and Δy displacements and empirically calculated cumulative density functions (**Supplementary Figure S1**, top left panel). These distributions were found to show slight deviations from a Gaussian distribution.

A Stochastic Model for Cell Migration

Our preliminary cell track analysis led us to model cell movement as a 2D random walk with persistence (**Figure 1C**). In our model, for each time interval i, the distance, r_i, traveled by a cell and the angle, θ_i, through which the cell moves are considered random variables. The random variable r_i is taken to have a Gaussian distribution characterized by mean μ_R, and variance σ_R^2. We allowed for negative values of r_i to account for the scenario in which a cell maintains its direction of polarization, but its centroid moves in a rearward direction. The directional angle θ_i, is also taken to have a Gaussian distribution with variance σ_s^2, and centered on the value of the previous angle θ_{i-1}. Small values of σ_s^2 correspond to highly persistent migration. For large values of σ_s^2 the new direction becomes uniformly distributed on the interval $[-\pi, \pi]$ and the model represents a purely diffusive process.

It is not possible to tell from cell track data alone if changes in θ_i of magnitude greater than $\pi/2$ resulted from large deviations in orientation or negative r_i. Thus, the probability distribution for these variables cannot be constructed unambiguously from the cell track data. To overcome this difficulty, we performed a change of variables from (r_i,θ_i) to $(\Delta x_i^{\parallel}, \Delta y_i^{\perp})$, where these new variables correspond to changes in the centroid's position during the *i*th time interval that are parallel and perpendicular to the direction of the previous step (**Figure 1C**). An important feature of the model is that analytical expressions for the probability density functions (pdfs) of $\Delta x_i^{\parallel}$ and $\Delta y_i^{\perp}$ can be found (**Supplementary Information**), allowing estimation of model parameters from experimental data to be performed in a computationally efficient manner, relative to the alternative of estimating probability density functions via repeated simulation of the stochastic model (**Figure 1C**). These co-ordinates explicitly handle the degeneracy in θ_i described above, because in these co-ordinates all possibilities that could have led to a given observation are considered. If cells show persistent motion, $\Delta x_i^{\parallel}$ has a positive mean value. Also, if there are no external cues in the experiments to define a preferred direction of motion, $\Delta y_i^{\perp}$ is symmetric about zero. Therefore, the distribution for $\Delta x_i^{\parallel}$ is more informative, and we use it to compare the experimental results with the model's behavior. It is possible to simultaneously fit the $\Delta x_i^{\parallel}$ and $\Delta y_i^{\perp}$ distributions, but this comes at an increased computational cost. As a consistency check, after performing parameter estimation, we verify that the model accurately captures the $\Delta y_i^{\perp}$ distribution. If the model failed this consistency test, we could repeat the parameter estimation using both distributions. However, this was not required for any of the cases considered here.

We used a Monte Carlo method based on the Metropolis algorithm to perform parameter estimation. This was followed by local optimization algorithms to identify parameters associated with the global minimum error between the model and data (**Supplementary Information**). To test the accuracy and

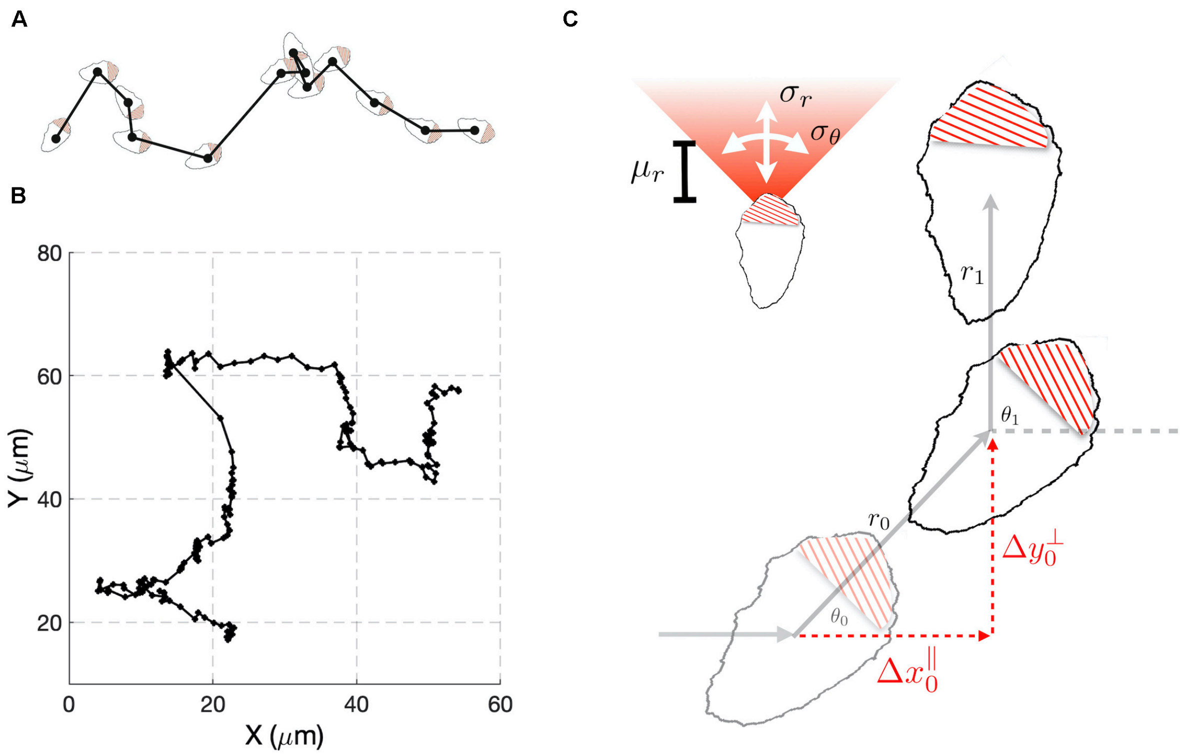


FIGURE 1 | A stochastic model for cell migration. **(A)** Cell tracks are constructed by recording the geometric center of the cell over time. **(B)** Example track resulting from tracking the cell centroid at 5 min intervals (black dots). **(C)** A stochastic model of migration in which during each time interval, a cell moves a distance *r* through and angle θ with respect to the direction of the previous step. The random variable *r* is taken to be normally distributed with mean μ_r and variance σ_r^2. The angle θ is also normally distributed with mean zero and variance σ_θ^2. To compare the model to experimental data we change variables to $\Delta x^{\parallel}$ and $\Delta y^{\perp}$, the directions parallel and perpendicular to previous step.

efficiency of this method, we benchmarked our approach using data generated from computational simulations of the stochastic model (**Supplementary Figure S2**). Having validated our computational methods, we next fit the model to the experimentally measured distributions. The model did not generate a good fit to experimental data for MEF cells (**Supplementary Figure S3**, dashed curve). In particular, we found that the model could not capture the second mode observed in the $\Delta x_i^{\parallel}$.

A Multistate Model for Cell Migration

Further inspection of the MEF cell tracks suggested that individual cells might exist in different modes of migration, distinguished by differences in speed and persistence. We therefore expanded our model to allow for different states of migration. That is, we hypothesized that at any given time a migrating cell is in one of n states denoted by S_i, with $i \in \{1 \ldots n\}$. Each state is characterized by the parameters μ_r^i, σ_r^i, and σ_θ^i. The additional parameters, α^i, denoting the fraction of time spent in state *i*, are required to fully specify the model. Since $\sum \alpha^i = 1$, in the two-state case the total number of parameters is seven. Note that if a two-state model is fit to data consisting of only a single state, then we expect our Monte Carlo method to produce parameter sets in which α^1 takes on values of 0 or 1, or $\mu_r^1 = \mu_r^2$, $\sigma_r^1 = \sigma_r^2$, and $\sigma_\theta^1 = \sigma_\theta^2$. The extended model is essentially a mixture model, which is itself a reduced hidden Markov model under the assumption that the probabilities of transitioning between states are independent and identically distributed. We again used simulated data to validate the accuracy and efficiency of our Monte Carlo method when multiple states are considered (**Supplementary Figure S4**).

The multi-state model produced a good fit to the MEF $\Delta x_i^{\parallel}$ distribution (**Figure 2A**). To assess the accuracy of our parameter estimates we used confidence-interval profiling (Raue et al., 2009). To determine acceptable values for the sum of the squared errors (SSE) we boot-strapped the original datasets to assess plausible differences in our observed distributions should we repeat the experiments (**Supplementary Information**). The results of this analysis provide a measure of the confidence that should be placed on each estimated parameter value (**Supplementary Figures S5A,B**). Of particular interest is the parameter α which represents the fraction of time in each state. The best fits were achieved with $\alpha = 0.12$. We confirmed that the model also captured the distributions for $\Delta y_i^{\perp}$ (**Supplementary Figure S6A**). The results of our analysis suggest that randomly migrating MEFs exist in one of two states. About 12% of time these cells are in a state with a well-defined characteristic step of ∼3 μm (State 1 – blue distribution in **Figure 2A** left

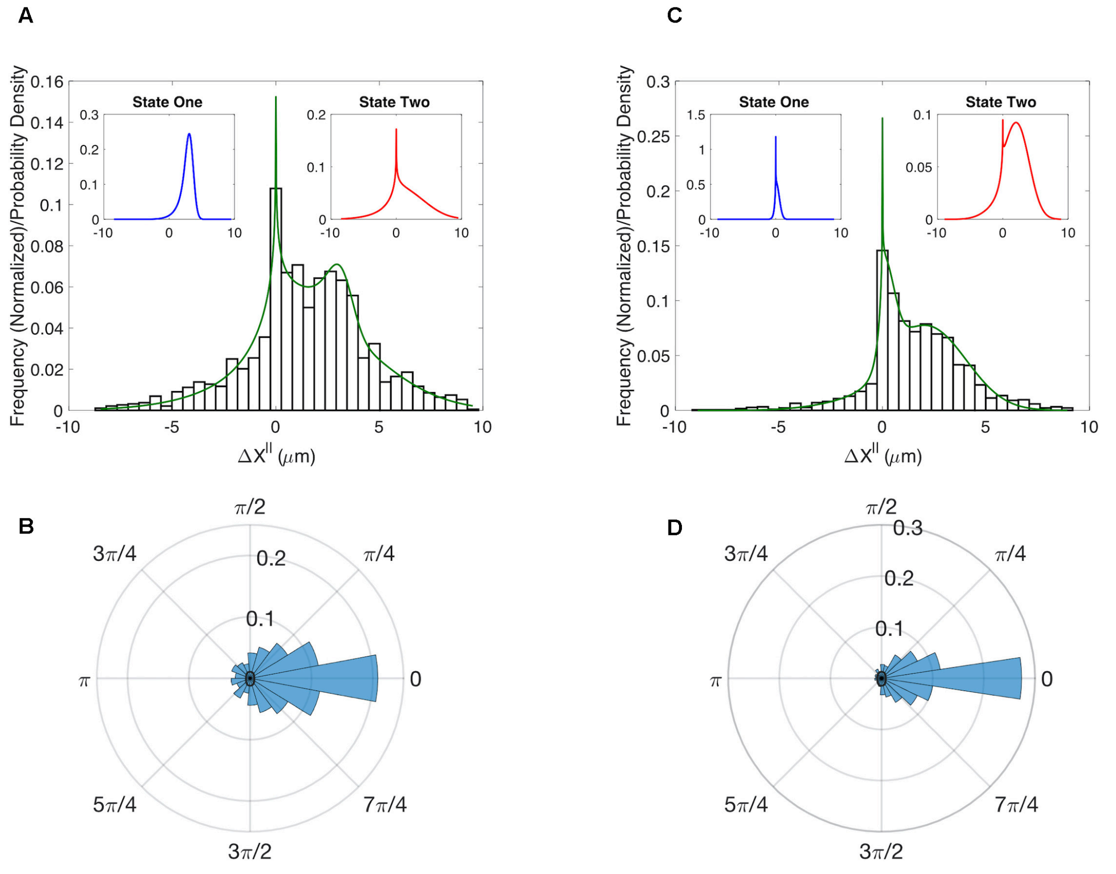


FIGURE 2 | Results for the multistate model of migration. **(A)** Comparison of the experimentally determined distribution of $\Delta x^{\|}$ for WT MEF cells (histogram) to the results of a two-state model (green curve). Insets show the distributions for the predicted two states. **(B)** Experimentally determined distribution for the angle θ for WT MEF cells. **(C)** Same as **(A)** except for cells in which RhoG has been knocked down. **(D)** Same as B except for cells in which RhoG has been knocked down.

inset) and an angular distribution with $\sigma_\theta^1 = 0.7$. In the second state, the step size is highly variable (State 2 – red distribution **Figure 2A** right inset) and the motion is less persistent $\sigma_\theta^2 = 1.3$. For completeness, we also show the distribution for the angle θ (**Figure 2B**).

RhoG's Role in Migration

It has long been appreciated that the canonical Rho-GTPases RhoA, Rac1, and Cdc42 play important roles in cell migration. However, the role of RhoG in migration is less well studied. To determine if RhoG plays a role in the random migration of MEFs, we generated time series data for cells in which this protein was knocked down. While the angular distributions for the WT and knockdown do not show clear differences (**Figures 2B,D**), the $\Delta x_{\|}$ distributions indicate RhoG does effect migration (**Figures 2A,C**). Moreover, by fitting our simple two-state model we can quantify this effect and ascertain that the persistent state 1 in the MEF control has been converted to a state in which the cells do not show significant movement ($\mu_r^1 = 3.2$ μm for the WT to $\mu_r^1 = 0.25$ μm for the KD). State 2 seems to be preserved by the KD in the sense that the confidence intervals defining state 2 parameters are overlapping in the two cases (**Supplementary Figure S5**). A putative mechanism for how RhoG activation influences cell migration via recruitment of the DOCK180/ELMO complex (Katoh and Negishi, 2003; Katoh et al., 2006), which acts as GEF for Rac1. However, whether this is the key pathway in this process, and how it is organized spatio-temporally, is a direction of future research.

Inferring States From Time Series Data

We next sought to develop computational tools that could be used to determine if the predicted states of migration correspond to subpopulations of cells with distinct phenotypes or if individual cells could transition between states. To test if individual cells change their migration state, we developed a method to infer migration states from individual cell tracks. Our approach uses a Bayesian prediction method based on the probability that a sequence of *k* successive steps arises from one of the identified states (see Methods for details). Before applying our state prediction method on the experimental data, we first validated the approach using synthetic data. To generate this data, we performed computational simulations of the stochastic model using the parameters estimated from the experimental data for MEF cells. With these values our state-prediction algorithm correctly identified the states more than 90% of the time, validating the approach (**Supplementary Figure S7**).

Having demonstrated our method's ability to infer cell migration states from simulated track data and demonstrate a role for RhoG, we examined whether the different migration states could be correlated with molecular changes within cells. The Hahn lab has used HeLa cells to develop new biosensors and optogenetic probes. It is well established that these molecular tools must be used at controlled concentrations, below levels where they perturb cell movement (Kraynov et al., 2000; Machacek et al., 2009). Controls in earlier studies have shown that HeLa cells exhibit altered motility when the Rac1 biosensor is expressed at high levels. We decided to investigate if our stochastic modeling approach could quantify the effects of biosensor overexpression. We compared WT HeLa cells without biosensor expression to cells with high levels of Rac1 biosensor. Our analysis revealed that WT cells showed little directed motion and a single migration state was sufficient to capture the distributions of steps sizes (**Figures 3A,B**). In contrast, cells with highest levels of biosensor exhibited two states of migration (**Figure 3C**). In particular, two states were needed to capture the long tail of the distribution (see **Supplementary Figure S9** for comparison of one-and two state results). In state 1 the cell moves in persistent manner, whereas in state 2 the cell is mostly stationary. To test if the predicted two states are correlated with differences in cell signaling, we ran our state prediction algorithm on the track data. Interestingly, our analysis predicted that individual cells randomly switch between the two states (**Figure 4A**), and qualitative observations indicated that the slow state showed multiple disperse Rac activation events at the edge of the cell, while the fast state showed a single Rac activation at the leading edge (see **Supplementary Movies M1, M2**). To quantify this observation, we identified and counted the number of foci of active Rac1 in each image and grouped these counts by the predicted state (see **Supplementary Information** for details), reasoning that random movement would require more cell protrusions distributed around the cell perimeter. Rac activation is known to be sufficient to generate cell protrusions (Wu et al., 2009; Wang et al., 2016). Cells predicted be in state 1, which corresponds to the fast-persistent state, had fewer Rac1 foci than

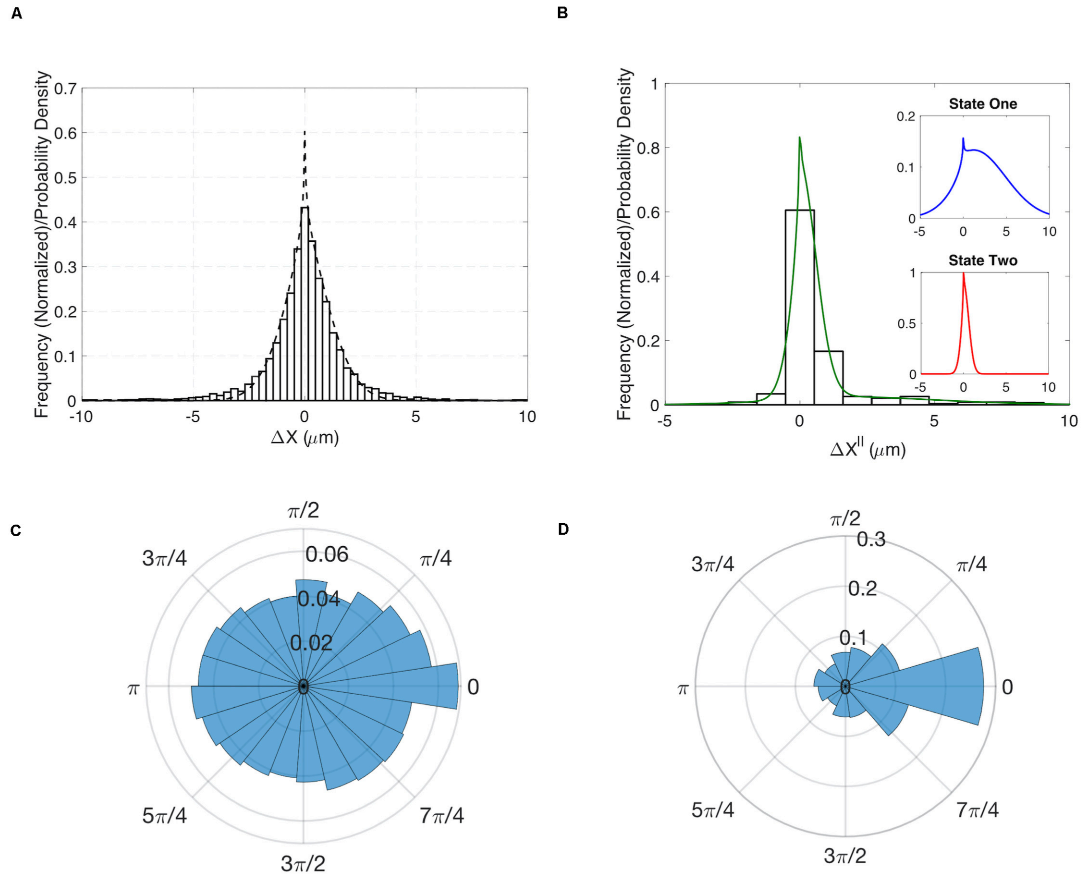


FIGURE 3 | Results for HeLa cells with and without expression of a Rac1 biosensor. **(A)** Comparison of the experimentally determined distribution of $\Delta x^{\|}$ for HeLa cells not expressing the biosensor (histogram) to the results of a one-state model (dashed curve). Insets show the distributions for the predicted two states. **(B)** Experimentally determined distribution for the angle θ for HeLa cells not expressing the biosensor. **(C)** Comparison of the experimentally determined distribution of $\Delta x^{\|}$ for HeLa cells expressing Rac1 biosensor (histogram) to the results of a two-state model (green curve). Insets show the distributions for the predicted two states. **(D)** Experimentally determined distribution for the angle θ for HeLa cells expressing biosensor.

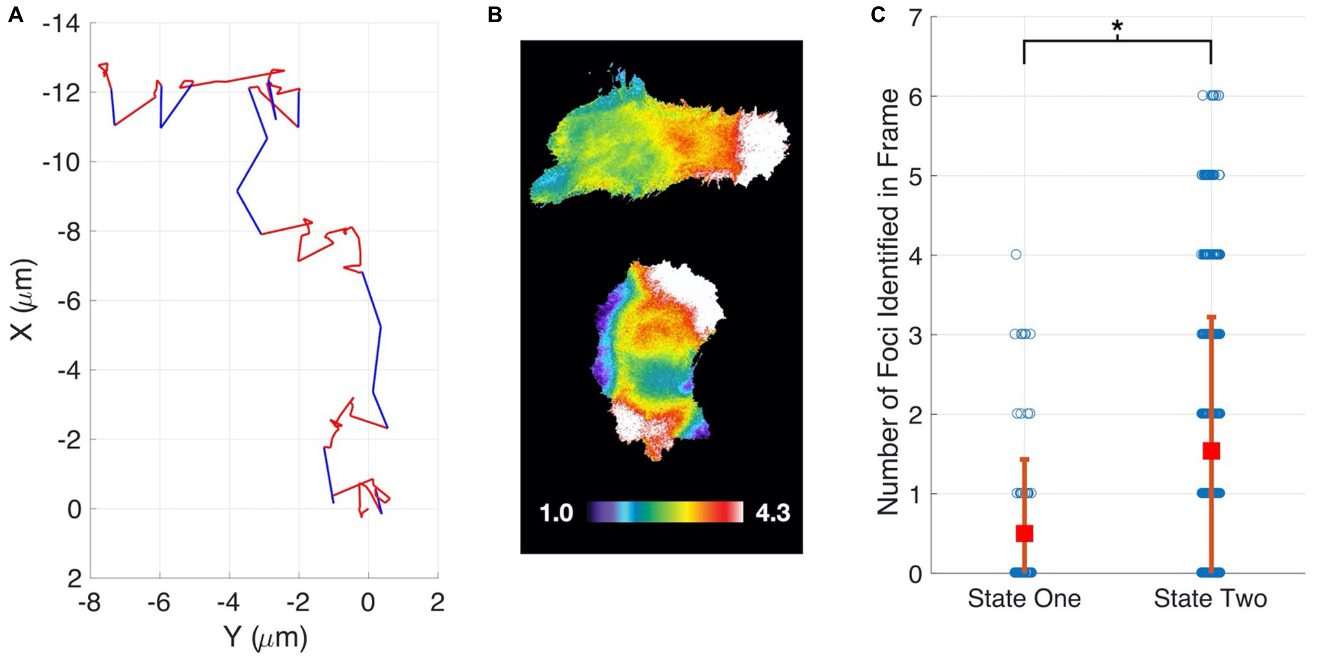


FIGURE 4 | State prediction for HeLa Cells expressing a Rac1 biosensor. **(A)** Example cell track showing switching between a fast-persistent state (blue) and slower-less persistent state (red). **(B)** When in the fast-persistent state (state 1, upper panel) cells have fewer active Rac1 foci as compared to the slower-less persistent state (state 2, lower panel). The upper cell is undergoing persistent motion to the right. The color scale indicates the ratiometric readout of Rac activation, normalized so the lowest 5% of the cell = 1. **(C)** Quantification of the number of foci in each state.

those predicted to be in state 2, which show little net movement (**Figure 4B**). This observation is consistent with highly motile cells typically showing strong polarity.

DISCUSSION

We developed novel computational methods for analyzing the movement of randomly migrating cells. Our approach combines stochastic modeling with statistical inference methods to detect and quantify migratory phenotypes. Migrating cells have a biochemical, morphological, and structural orientation that persists as these cells move. Our model captures this 'memory' by conditioning the cell's movement during the current time interval on its previous direction of motion. An important feature of our model is that analytic expressions for the probability densities for cell displacements parallel and perpendicular to the previous direction of motion can be found. This feature allows us to generate the probability density function for a given set of parameters rather than generating an approximation to this PDF via stochastic simulation of our migration model (**Figure 1C**). In most use cases the analytical PDF is computationally more efficient due to the high number of repeats required to estimate the PDF with suffcient accuracy. We have validated all our approaches using simulated data, and then applied the methodology to study randomly migrating MEF and HeLa cells.

Our modeling approach allows for multiple states of migration. This feature allowed us to demonstrate that migrating cells randomly transition between modes of movement. Crucial to the detection of these states is the quantification of parameter values and the associated confidence in those estimates. This process allowed us to be confident in the existence of two states of migration for MEF and HeLa cells over-expressing a Rac1 biosensor.

The identification of multiple states of migration for MEF cells led us to assess the role of RhoG in establishing these states. To do this we used siRNA to reduce RhoG expression. This perturbation suggests that RhoG plays a role in directed migration, because reducing RhoG eliminated net movement in the first predicted state and shortened the range of step sizes taken in the second state. We next developed a Bayesian approach to predict the current migration state of a cell from time series of the cell's position. Using this method, we demonstrated that individual HeLa cells expressing a Rac1 biosensor switched between migratory states. Importantly, we were able to correlate these two states with differences in the distribution of Rac1 activity.

We believe that our methods provide useful tools for quantifying and characterizing cell migration. Our stochastic model characterizes cell migration using parameters with straightforward biological interpretations. Hence, application of this model can lead to biological insights not apparent in the data from visual inspection or simple quantitative measures. In this case, our analysis suggests a role of RhoG in allowing cells to change direction, which may play a role in the ability of randomly migrating cells to search their environment.

MATERIALS AND METHODS

Computational Methods

The full code and analysis for this paper is available (Allen, 2020).

Coordinate Transformation

We modeled cell migration as a stochastic sequence of steps characterized by the step size r_i and directional angle θ_i (**Figure 1**). Since we assume r_i and θ_i to be realizations of independent random variables R and Θ the probability the cell moves (r, θ) is defined by

$$f(r, \theta|\theta_{i-1}) = g_R(r).g_\Theta(\theta|\theta_{i-1}) \tag{1}$$

where $g_R(r)$ is the probability density function (pdf) for the step magnitude, which we take to have the normal distribution $\mathcal{N}(\mu_r, \sigma_r^2 \Delta t)$, and $g_\Theta(\theta|\theta_{i-1})$ is the pdf generating the new orientation conditioned on the previous angle, which we take to have the normal distribution $\mathcal{N}(\theta_{i-1}, \sigma_\theta^2 \Delta t)$. The experimental data is collected in Cartesian coordinates *(X, Y)*. In principle we could transform the data into the coordinates R and Θ. However this transformation cannot be completed uniquely, because there is no way to distinguish a backward step in which the cell maintains its direction of polarity ($\theta_i = \theta_{i-1}$) from one in which the front and back of the cell have reversed ($\theta_i = \theta_{i-1} + (2k + 1)\pi$). Furthermore, the value of θ_i cannot be determined if $r_i = 0$. For these reasons, we transform the model to the coordinates $(\Delta x_i^{\parallel}, \Delta y_i^{\perp})$, where these new variables correspond to changes in the centroid's position during the ith time interval that are parallel and perpendicular to the direction of the previous step.

To compare with the model the data needs to be manipulated to generate histograms for steps in the $x^{\parallel}$ and $y^{\perp}$ directions. For each sequential triplet of coordinates $\{(x_{i-1}, y_{i-1}), (x_i, y_i), (x_{i+1}, y_{i+1})\}$, we rotate the steps as a rigid body about (x_{i-1}, y_{i-1}) by a four quadrant inverse tangent based on $\tan^{-1}\frac{y_i - y_{i-1}}{x_i - x_{i-1}}$. The result of this is that all steps are pre-orientated in a positive x-direction and initiated at $(0, 0)$, and can be plotted as histograms of step distance in the x and y direction: $(\Delta x_i^{\parallel}, \Delta y_i^{\perp}) = (x'_{i+1} - x'_i, y'_{i+1} - y'_i)$.

The pdf for $\Delta x^{\parallel}$ is:

$$f_X(\Delta x^{\parallel}) = \int_{-1}^{1} \frac{f_H(\Delta x^{\parallel}/h, h)}{h} dh \tag{2}$$

where $h = \cos(\theta)$, and,

$$f_H(\Delta x^{\parallel}/h, h) = g_r(\Delta x^{\parallel}/h) \sum_k \frac{1}{(1-h^2)^{\frac{1}{2}}} \times \\ (g_\Theta(\arccos(\mathrm{h}) + 2\pi k) + g_\Theta(-\arccos(\mathrm{h}) + 2\pi k)). \tag{3}$$

The expression for $f_Y(\Delta y^{\perp})$ is similar, however now with $g_\Theta(\arccos(\mathrm{h}) + \pi/2 + 2\pi k) + g_\Theta(-\arccos(\mathrm{h}) + \pi/2 + 2\pi k)$ in the summation term. A derivation of these results is presented in the **Supplementary Information**.

Parameter Estimation

Parameters were estimated by simulated annealing, which is a Monte Carlo method based on the Metropolis algorithm (24, 25). Initial choices of parameters generate an analytical solution (Eq. 2), which is scored against the experimental data $(\Delta x_i^{\parallel})$ by the sum of least squared differences. At each step of the algorithm the parameters are updated by a small addition of Gaussian noise, if this update scores better than the current score then these parameters are accepted. If the score is higher, the parameter set is accepted with probability $e^{-\frac{\Delta s}{T}}$, where Δs is the difference between the current and previous scores and T is the current temperature. Over the course of the fitting T, the temperature is reduced. This fixes the parameter choices into a local minimum. Here we choose a geometric cooling regime. Due to the stochastic nature of the simulation, and that there could be many local minima, it is necessary to run this fitting procedure multiple times. The best fit of this routine was then further refined using MATLABs fmincon routine, which was also used to assess the sensitivity of our fit to altering parameter values via confidence-interval profiling (**Supplementary Figure S4**, **Supplementary Material** for details).

The histograms were amalgamated from multiple cell tracks. For the case of two states, the pdf for Δx becomes

$$f_X(\Delta x) = \alpha f_X^1(\Delta x) + (1 - \alpha) f_X^2(\Delta x)$$

where α is the fraction of time spent in state 1 and the distributions $f_X^1(\Delta x)$ and $f_X^2(\Delta x)$ are parameterized by $(\mu_r^1, \sigma_r^1, \sigma_\theta^1)$ and $(\mu_r^2, \sigma_r^2, \sigma_\theta^2)$, respectively.

Parameter sets were identified by multiple simulated annealing runs, followed by local-optimization routines.

Validation of Methods

To validate the pdfs and the parameter estimation algorithm, we simulated cell tracks using the stochastic model (**Figure 1**). Cell tracks were generated using two states, each with distinct parameter sets. At each step a state was chosen at random with probability 0.5. As above, the simulated cell tracks were used to construct the distributions for $\Delta x_i^{\parallel}$ and $\Delta y_i^{\perp}$. We assumed model parameters were not known and used the Monte Carlo method to fit Eq. 2, modified to two states (see below) to the simulated data for $\Delta x_i^{\parallel}$. The Monte Carlo method quickly converged on the correct parameter values (**Supplementary Figure S4**), validating the analytical solution to the model and our fitting procedure. In theory we also could fit the pdf for $\Delta y_i^{\perp}$. However, the pdf for $\Delta y_i^{\perp}$ is symmetric, because there is no preferred direction of migration and therefore less informative than the distribution for $\Delta x_i^{\parallel}$. We found that we could maintain the accuracy of our parameter estimation while improving the computational cost by only considering the $\Delta x_i^{\parallel}$ distribution. As a consistency check, we always verify that the estimated parameters accurately reproduce the pdfs for $\Delta y_i^{\perp}$ (**Supplementary Figure S6**).

State Prediction

To identify which state a cell is in at a given time, we used Bayes' theorem to invert the problem. That is, we calculate the probability that a cell is in state S_i given the experimental data. Note that in calculating this probability, we also get the false positive rate or p-value. To make a reliable prediction of

S_i may require an n-step window, where n is odd, such that, $\{x_{i-n/2}, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{i+n/2}\}$. Then:

$$P(s_i|X) = \frac{P(X|s_i)P(s_i)}{P(X|s_i)P(s_i) + P(X|s_i^c)P(s_i^c)}$$

where $P(X|s_i)$ is calculated from the model, and we take $P(s_i) = \alpha$. Windows of length one, three and five were tested. For the case presented here, we found that the window of length one produced results similar to the other two window lengths.

Foci Identification

Ratiometric images of the FRET based Rac1 biosensor were analyzed for localized regions of higher Rac1 activity near the periphery of the cell. We call these regions "foci". We used custom application of the image processing toolbox in MATLAB to identify foci, which we define as contiguous regions within the cell that were simultaneously: (1) 60% above the average intensity of the cell, (2) greater than 100 pixels in area, and (3) contained at least one point within 5 pixels of the cell edge. The length of time (or number of frames) that a cell could be followed for varied. So, to not overweight any one cell, the number of image frames analyzed, n, was selected to maximize $n \times m$ where m is the number of cells with at least n images.

Experimental Methods

Cell Culture and Transient Transfections

HeLa cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) (Cellgro) supplemented with 10% fetal bovine serum (FBS) (HyClone), 100 U/mL penicillin and 100 μg/mL streptomycin (Cellgro) and 2 mM L-glutamine (Invitrogen) at 37°C and 5% CO_2. All cDNA constructs were transfected into cells using FuGene6 (Roche) according to the manufacturer's instructions. IA32 Mouse Embryonic Fibroblast (MEF) cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) (Cellgro) supplemented with 10% fetal bovine serum (FBS) (HyClone) and 1× GlutaMAX (Thermo Fisher Scientific).

IA32MEFs were transfected with either RhoG siRNA (CAGGTTTACCTAAGAGGCCAA) or Allstars Negative Control siRNA (Qiagen, United States). 7.5 μL, 10 μM siRNA was added to 250 μL serum-free DMEM. 3 μL lipofectamine RNAimax was added to another 250 μL serum-free DMEM. After 5 min, the two solutions were mixed and incubated for 20 min, followed by dropwise addition to a 35 mm dish. Medium was changed after 24 h and cells were split as required for use in experiments 48–72 h post-transfection, when knock-down efficiency was maximal. Control siRNA cells were incubated with 5 μM CFDA green for 20 min in serum-free DMEM. CFDA-labeled control cells were mixed with unlabeled RhoG siRNA cells immediately prior to the experiment.

Live Cell Imaging

For live cell imaging, cells were plated on fibronectin-coated coverslips (10 μg/ml fibronectin) 4 h before imaging, then transferred to Ham's F12-K imaging medium supplemented with 2% FBS and 15 mM HEPES. Live cell imaging was performed in a closed heated chamber (20/20 Bionomic).

For biosensor imaging, photobleach-corrected time-lapse image stacks were acquired for 18 h at 5 min intervals and processed as previously described (Pertz et al., 2006; Machacek et al., 2009; Hodgson et al., 2010). The following filter sets were used (Chroma Technology Corp.): CFP: D436/20, D470/40; FRET: D436/20, ET535/30; YFP: D500/20, ET535/30. Cells were illuminated with a 100 W Hg arc lamp through a 1.0 neutral density filter.

For RhoG siRNA experiments cell tracks were generated through 10× DIC imaging of cells plated as above, but using Ham's F12K medium supplemented with 5% FBS. Images were acquired for at least 70 frames at 10 min intervals in a closed, heated chamber. This length of track was objectively identified as optimal by maximizing the total number of analyzed frames in the entire data set.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RA, TE, and KH devised the research plan and wrote the manuscript. RA performed the mathematical and computational calculations. CW and NP collected the data. All authors contributed to the article and approved the submitted version.

FUNDING

The work was supported by grants R35GM127145 (TE), R35GM122596 (KH), and W911NF-15-1-063 (TE and KH). A previous version of this manuscript has been released as a pre-print at BioRxiv (Allen et al., 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00822/full#supplementary-material

FIGURE S1 | Cumulative distribution functions (CDFs) for *x* and *y* step sizes compared to the CDFs for normal distributions.

FIGURE S2 | Validation of analytical approach. Simulated data (histogram) is generated by simulating data (blue bars) stochastically with parameters $(\mu_r, \sigma_r, \sigma_\theta, \sigma_0) = (0.1, 0.1, 0.5, \pi/5)$. Comparison with the analytical PDF (red) gives good agreement.

FIGURE S3 | Fitting the model to MEF WT cells indicates that a one state model of migration cannot adequately capture key features of the data. Observed data (open bars), model fit (green).

FIGURE S4 | Analytics and Fitting Validation. Simulated data (histogram) is generated by applying the model with parameters $(\mu_r^1, \sigma_r^1, \sigma_\theta^1) = (1, 0.3, 3)$, $(\mu_r^2, \sigma_r^2, \sigma_\theta^2) = (5, 1, 0.5)$ and $\alpha = 0.5$. Assuming the underlying parameters were unknown, we used simulated annealing to fit the analytical solution (green line). This validates our analytical solution and fitting procedure (we estimate $(\hat{\mu}_r^1, \hat{\sigma}_r^1, \hat{\sigma}_\theta^1) = (1.02, 0.30, 2.72)$, $(\hat{\mu}_r^2, \hat{\sigma}_r^2, \hat{\sigma}_\theta^2) = (5.08, 0.93, 0.51)$, and $(\alpha = 0.50)$.

FIGURE S5 | Parameter values and associated confidence intervals. For each condition, and each parameter, the best-fit is the minimum of the blue curve. Acceptable parameter values are those which can lead to a value below the threshold (red dots) when the other parameters are re-fit, highlighted by the red-dash lined. Note, that two states were clearly identified in both cases.

FIGURE S6 | Model prediction versus the ΔY distribution from the cases shown in **Figure 2**. **(A)** MEF WT **(B)** MEF RhoG KD. In **(A)** and **(B)**: Observed data (open bars), model fit (green) and individual pdfs for state one and state two (insets).

FIGURE S7 | Testing the accuracy of predicting states using HeLa Rac1 model parameters. Data simulated stochastically, where we know the state a given step was generated from, is compared with our prediction of the state using Bayes Theorem. Overall, we are correct more than 90% of the time. However, steps truly in state one get occasionally mischaracterized are overlapping leading to an accuracy of around 70% in this case.

FIGURE S8 | Bootstrapped Analysis of the Difference in Foci Between States. To ascertain if the difference we observed in the mean foci count in state one versus state two was meaningful or could have been observed by chance, we performed a boostrap analysis. For a given iteration of the bootstrap, we randomly assigned state one or state two to each image frame with probability α and $1 - \alpha$, respectively. Then, for each iteration, calculated Δ mean foci as the difference in the mean number of foci in state one versus state two. Blue bars, bootstrapped distribution (50,000 iterations), dotted red line observed Δ mean foci.

FIGURE S9 | Comparison of 1-state (dashed curves) and 2-state (green curves) model fits for all of the experimental results.

FIGURE S10 | Overview of algorithm to estimate model parameters and confidence intervals.

FILES M1 | M1 (slow_state_hela_rac1.mp4) and M2 (fast_state_hela_rac1.mp4). The predicted "fast state" of migration was correlated with fewer Rac1 foci, most frequently a single focus at the leading edge, and in the slow state multiple brief Rac1 activation foci appear.

REFERENCES

Allen, R. J. (2020). *Cell Tracking Project*. Available at: https://github.com/rallen81/CellTracking (accessed January 17, 2018).

Allen, R. J., Welch, C., Pankow, N., Hahn, K., and Elston, T. C. (2018). Stochastic methods for inferring states of cell migration. *bioRxiv [Preprint]*. doi: 10.1101/249656

Arrieumerlou, C., and Meyer, T. (2005). A local coupling model and compass parameter for eukaryotic chemotaxis. *Dev. Cell* 8, 215–227. doi: 10.1016/j.devcel.2004.12.007

Cain, R. J., and Ridley, A. J. (2012). Phosphoinositide 3-kinases in cell migration. *Biol. Cell* 101, 13–29. doi: 10.1042/bc20080079

Dieterich, P., Klages, R., Preuss, R., and Schwab, A. (2008). Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. U.S.A.* 105, 459–463. doi: 10.1073/pnas.0707603105

Dimilla, P. A., Quinn, J. A., Albelda, S. M., and Lauffenburger, D. A. (1992). Measurement of individual cell migration parameters for human tissue cells. *AIChE J.* 38, 1092–1104. doi: 10.1002/aic.690380712

Finney, A. C., Stokes, K. Y., Pattillo, C. B., and Orr, A. W. (2017). Integrin signaling in atherosclerosis. *Cell Mol. Life Sci.* 74, 2263–2282. doi: 10.1007/s00018-017-2490-4

Franca-Koh, J., Kamimura, Y., and Devreotes, P. N. (2007). Leading-edge research: PtdIns(3,4,5)P3 and directed migration. *Nat. Cell Biol.* 15–17. doi: 10.1038/ncb0107-15

Goley, E. D., and Welch, M. D. (2006). The ARP2/3 complex: an actin nucleator comes of age. *Nat. Rev. Mol. Cell Biol.* 7, 713–726. doi: 10.1038/nrm2026

Hall, A. (2009). The cytoskeleton and cancer. *Cancer Metastasis Rev.* 28, 5–14.

Hodgson, L., Shen, F., and Hahn, K. (2010). Biosensors for characterizing the dynamics of rho family GTPases in living cells. *Curr. Protoc. Cell Biol.* 14, Unit 14.11.1–26. doi: 10.1002/0471143030.cb1411s46

Iden, S., and Collard, J. G. (2008). Crosstalk between small GTPases and polarity proteins in cell polarization. *Nat. Rev. Mol. Cell Biol.* 9, 846–859. doi: 10.1038/nrm2521

Jaffe, A. B., and Hall, A. (2005). "Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.* 21, 247–269.

Katoh, H., Hiramoto, K., and Negishi, M. (2006). Activation of Rac1 by RhoG regulates cell migration. *J. Cell Sci.* 119, 56–65. doi: 10.1242/jcs.02720

Katoh, H., and Negishi, M. (2003). RhoG activates Rac1 by direct interaction with the Dock180-binding protein Elmo. *Nature* 424, 461–464. doi: 10.1038/nature01817

Kraynov, V. S., Chamberlain, C., Bokoch, G. M., Schwartz, M. A., Slabaugh, S., and Hahn, K. M. (2000). Localized Rac activation dynamics visualized in living cells. *Science* 290, 333–337. doi: 10.1126/science.290.5490.333

Ladwein, M., and Rottner, K. (2008). On the Rho'd: the regulation of membrane protrusions by Rho-GTPases. *FEBS Lett.* 582, 2066–2074. doi: 10.1016/j.febslet.2008.04.033

Lemarié, C. A., Tharaux, P. L., and Lehoux, S. (2010). Extracellular matrix alterations in hypertensive vascular remodeling. *J. Mol. Cell. Cardiol.* 48, 433–439. doi: 10.1016/j.yjmcc.2009.09.018

Machacek, M., Hodgson, L., Welch, C., Elliott, H., Pertz, O., Nalbant, P., et al. (2009). Coordination of Rho GTPase activities during cell protrusion. *Nature* 461, 99–103. doi: 10.1038/nature08242

Marston, D. J., Anderson, K. L., Swift, M. F., Rougie, M., Page, C., Hahn, K. M., et al. (2019). High Rac1 activity is functionally translated into cytosolic structures with unique nanoscale cytoskeletal architecture. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1267–1272. doi: 10.1073/pnas.1808830116

Othmer, H. G., Dunbar, S. R., and Alt, W. (1988). Models of dispersal in biological systems. *J. Math. Biol.* 26, 263–298. doi: 10.1007/bf00277392

Pertz, O., Hodgson, L., Klemke, R. L., and Hahn, K. M. (2006). Spatiotemporal dynamics of RhoA activity in migrating cells. *Nature* 440, 1069–1072. doi: 10.1038/nature04665

Petrie, R. J., Doyle, A. D., and Yamada, K. M. (2009). Random versus directionally persistent cell migration. *Nat. Rev. Mol. Cell Biol.* 10, 538–549. doi: 10.1038/nrm2729

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., et al. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25, 1923–1929. doi: 10.1093/bioinformatics/btp358

Ridley, A. J. (2006). Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking. *Trends Cell Biol.* 16, 522–529. doi: 10.1016/j.tcb.2006.08.006

Ridley, A. J. (2015). Rho GTPase signalling in cell migration. *Curr. Opin. Cell Biol.* 36, 103–112. doi: 10.1016/j.ceb.2015.08.005

Rosello, C., Ballet, P., Planus, E., and Tracqui, P. (2004). Model driven quantification of individual and collective cell migration. *Acta Biotheor.* 52, 343–363. doi: 10.1023/b:acbi.0000046602.58202.5e

Rottner, K., Hall, A., and Small, J. V. (1999). Interplay between rac and rho in the control of substrate dynamics. *Curr. Biol.* 9, 640–648.

Schaks, M., Giannone, G., and Rottner, K. (2019). Actin dynamics in cell migration. *Essays. Biochem.* 63, 483–495.

Svensson, C. M., Medyukhina, A., Belyaev, I., Al Zaben, N., and Figge, M. T. (2018). Untangling cell tracks: quantifying cell migration by time lapse image data analysis. *Cytometry* 93, 357–370. doi: 10.1002/cyto.a.23249

Wang, H., Vilela, M., Winkler, A., Tarnawski, M., Schlichting, I., Yumerefendi, H., et al. (2016). LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat. Meth.* 13, 755–758. doi: 10.1038/nmeth.3926

Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B., et al. (2009). A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* 461, 104–108. doi: 10.1038/nature08241

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 28 July 2020
doi: 10.3389/fphys.2020.00896

A Quantitative Modular Modeling Approach Reveals the Effects of Different A20 Feedback Implementations for the NF-kB Signaling Dynamics

Janina Mothes[1], Inbal Ipenberg[2†], Seda Çöl Arslan[2], Uwe Benary[1], Claus Scheidereit[2] and Jana Wolf[1*]

[1] Mathematical Modelling of Cellular Processes, Max Delbrück Center for Molecular Medicine, Berlin, Germany, [2] Signal Transduction in Tumor Cells, Max Delbrück Center for Molecular Medicine, Berlin, Germany

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
Mariko Okada-Hatakeyama,
Osaka University, Japan
Didier Gonze,
Université Libre de Bruxelles, Belgium

****Correspondence:***
Jana Wolf
jana.wolf@mdc-berlin.de

†Present address:
Inbal Ipenberg,
Institut für Neuroimmunologie und Multiple Sklerose, Zentrum für Molekulare Neurobiologie Hamburg, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *30 April 2020*
Accepted: *02 July 2020*
Published: *28 July 2020*

Citation:
Mothes J, Ipenberg I, Çöl Arslan S, Benary U, Scheidereit C and Wolf J (2020) A Quantitative Modular Modeling Approach Reveals the Effects of Different A20 Feedback Implementations for the NF-kB Signaling Dynamics. Front. Physiol. 11:896. doi: 10.3389/fphys.2020.00896

Signaling pathways involve complex molecular interactions and are controled by non-linear regulatory mechanisms. If details of regulatory mechanisms are not fully elucidated, they can be implemented by different, equally reasonable mathematical representations in computational models. The study presented here focusses on NF-κB signaling, which is regulated by negative feedbacks via IκBα and A20. A20 inhibits NF-κB activation indirectly through interference with proteins that transduce the signal from the TNF receptor complex to activate the IκB kinase (IKK) complex. A number of pathway models has been developed implementing the A20 effect in different ways. We here focus on the question how different A20 feedback implementations impact the dynamics of NF-κB. To this end, we develop a modular modeling approach that allows combining previously published A20 modules with a common pathway core module. The resulting models are fitted to a published comprehensive experimental data set and therefore show quantitatively comparable NF-κB dynamics. Based on defined measures for the initial and long-term behavior we analyze the effects of a wide range of changes in the A20 feedback strength, the IκBα feedback strength and the TNFα stimulation strength on NF-κB dynamics. This shows similarities between the models but also model-specific differences. In particular, the A20 feedback strength and the TNFα stimulation strength affect initial and long-term NF-κB concentrations differently in the analyzed models. We validated our model predictions experimentally by varying TNFα concentrations applied to HeLa cells. These time course data indicate that only one of the A20 feedback models appropriately describes the impact of A20 on the NF-κB dynamics in this cell type.

AUTHOR SUMMARY

Models are abstractions of reality and simplify a complex biological process to its essential components and regulations while preserving its particular spatial-temporal characteristics. Modeling of biological processes is based on assumptions, in part to implement the necessary simplifications but also to cope with missing knowledge and experimental information. In consequence, biological processes have been implemented

by different, equally reasonable mathematical representations in computational models. We here focus on the NF-κB signaling pathway and develop a modular modeling approach to investigate how different implementations of a negative feedback regulation impact the dynamical behavior of a computational model. Our analysis shows similarities of the models with different implementations but also reveals implementation-specific differences. The identified differences are used to design and perform informative experiments that elucidate unknown details of the regulatory feedback mechanism.

Keywords: quantitative modeling, interlocked feedback loops, regulation, NF-κB signaling, A20, IKK regulation, response time, signaling dynamics

INTRODUCTION

Transcription factor NF-κB regulates cell differentiation, proliferation, and survival. In line with its broad range of normal physiological functions, aberrant activation of NF-κB can lead to severe diseases, e.g., autoimmune, neurodegenerative, and cardiovascular diseases as well as cancer and diabetes (Hayden and Ghosh, 2012; Perkins, 2012). In resting cells, the transcription factor NF-κB is located in the cytoplasm bound to IκBα, which prevents the translocation of NF-κB into the nucleus. Upon stimulation, e.g., with TNFα, the IκB kinase (IKK) complex is activated. The IKK complex phosphorylates IκBα, marking it for proteasomal degradation. Released NF-κB translocates into the nucleus and activates the transcription of a number of target genes (Hinz and Scheidereit, 2014). Two of these are NFKBIA, encoding IκBα, and TNFAIP3, encoding A20. Both proteins exhibit negative feedbacks on NF-κB activation. IκBα binds to NF-κB retrieving it from the DNA and thus exhibiting a direct negative feedback (Huxford et al., 1998). A20 inhibits NF-κB activity indirectly through interference with proteins mediating the signal from the TNF receptor complex to the IKK complex (Lork et al., 2017). The exact molecular mechanism of the inhibitory effect of A20 on the IKK complex is still under discussion (Skaug et al., 2011; De et al., 2014; Wertz et al., 2015).

In the last decades, several mathematical models describing the NF-κB signaling in different cell lines have been published (Hoffmann et al., 2002; Lipniacki et al., 2004; Longo et al., 2013; Zambrano et al., 2014; Fagerlund et al., 2015; Mothes et al., 2015; Murakawa et al., 2015; Benary and Wolf, 2019), and reviewed (Lipniacki and Kimmel, 2007; Cheong et al., 2008; Basak et al., 2012; Williams et al., 2014). These models describe the transient NF-κB activation or the oscillatory dynamics observed experimentally. It was also studied which factors can lead to a switch between oscillatory and non-oscillatory NF-κB dynamics (Mothes et al., 2015). All models comprise the core processes of the canonical NF-κB signaling, e.g., the interaction of NF-κB and IκBα and the transcription and translation of IκBα as well as the IKK-induced degradation of IκBα. The majority of those models include only the negative feedback via IκBα, which has been well-studied and characterized (Fagerlund et al., 2015).

Until today, only a small number of mathematical models has been developed that include the A20-dependent negative feedback mechanism (Lipniacki et al., 2004; Werner et al., 2008; Ashall et al., 2009; Murakawa et al., 2015). These models utilize similar implementations of the core signaling processes but differ in their implementation of the A20 feedback. Since the exact inhibitory mechanism of A20 on IKK has not yet been fully elucidated and may also vary between cell lines, the models implement different mechanisms. While the model of Lipniacki et al. (2004) and the derived model by Ashall et al. (2009) implement the inhibitory action of A20 on the level of IKK, the models of Werner et al. (2008) and Murakawa et al. (2015) basically implement the hypothesis that A20 blocks the signaling upstream of IKK by binding to TNF receptor associated proteins. In particular, the models by Lipniacki et al. (2004) and Ashall et al. (2009) comprise three different states of IKK: neutral, active and inactive. In the model proposed by Lipniacki et al. (2004), A20 promotes the inactivation of activated IKK, whereas, in the model by Ashall et al. (2009) A20 inhibits the "recycling" of inactive IKK to neutral IKK and consequently the activation of IKK. In the models by Werner et al. (2008) and Murakawa et al. (2015), A20 inhibits basal and TNFα-induced IKK activation, although Werner et al. (2008) consider the signaling mechanisms upstream of IKK with substantially more molecular detail than Murakawa et al. (2015). In short, all four models share a feedback inhibition of IKK activity by A20 but differ in the specifics of their A20 feedback implementations.

Here, we ask whether these different A20 implementations have effects for the NF-κB dynamics. This knowledge is required when choosing an available published model for the description of a new data set. For our comparison we selected the different A20 feedback structures implemented in the models of Lipniacki et al. (2004), Ashall et al. (2009), and Murakawa et al. (2015), because these capture three different hypotheses and the models are comparable at their level of detailedness. In contrast, the model by Werner et al. (2008) is very detailed, including 38 parameters for the upstream part. We addressed the question whether the different feedback implementations affect NF-κB dynamics in similar or distinct ways. To this end, we used a computational approach in which we established three ordinary differential equation (ODE) models. Each model is composed of a core module and an upstream module (**Figure 1A**). The core module is identical

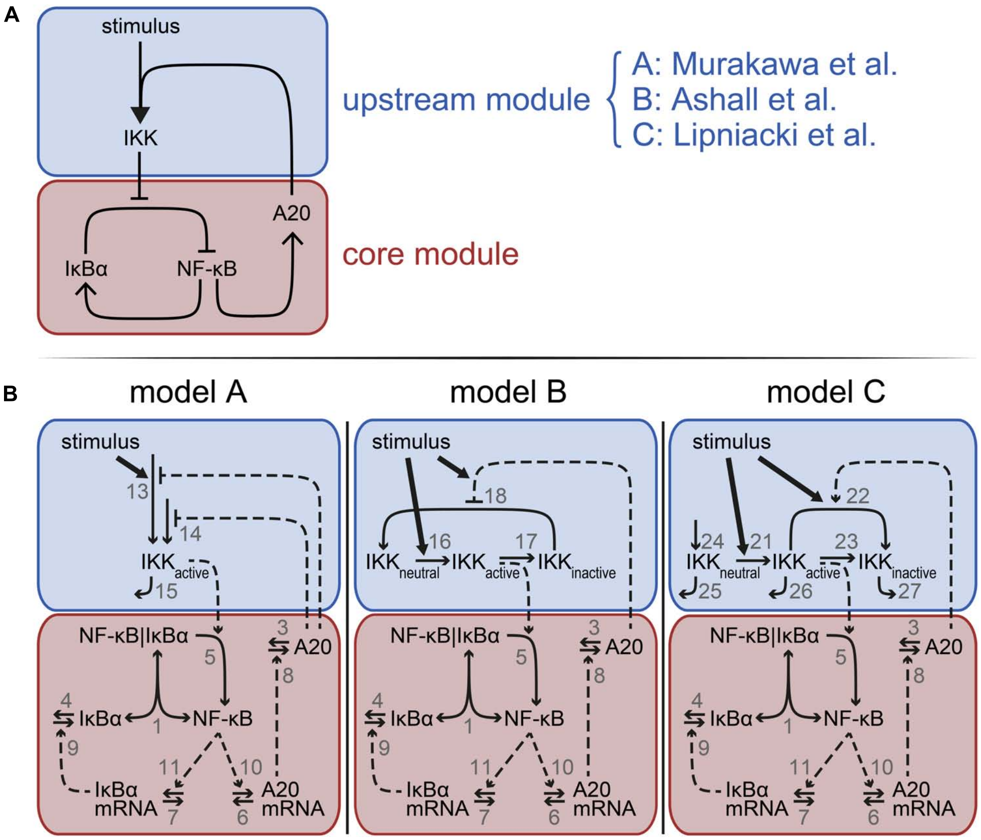


FIGURE 1 | Model schemes comprising the common core module and distinct upstream modules. **(A)** Each model is composed of a core module (red) and an upstream module (blue). The core module is identical in each model but the upstream module differs between model A, B, and C, implementing the A20 feedback mechanisms proposed by Lipniacki et al. (2004), Ashall et al. (2009), and Murakawa et al. (2015), respectively. **(B)** Schematic representations of the three models A–C. Vertical bars separate components in a complex. One-headed arrows indicate the direction of the reaction; double-headed arrows illustrate reversible binding reactions. Dashed arrows represent activation processes; the dashed lines ending in T-shape denote inhibition. The number next to an arrow specifies the number of the reaction. Model equations and the reference parameters are provided in the **Supplementary Information**.

in all three models and describes the interaction of NF-κB and IκBα, transcription and translation of IκBα, and IKK-induced degradation of IκBα. The three upstream modules comprise the three distinct mechanisms of IKK inhibition by A20 that Lipniacki et al. (2004), Ashall et al. (2009), and Murakawa et al. (2015) have proposed. In this way, we applied a modular concept to derive three models that share an identical core module but differ in their implementations of the A20 feedback in the upstream module. By fitting these models to a set of published experimental data, we derive three models showing quantitatively similar NF-κB dynamics. We use this computational approach to directly compare the influences of the structural difference in the upstream modules on the response of the NF-κB dynamics. In particular, we focused on the impact of the A20 and IκBα feedback strength. Moreover, we analyze in each model how the A20 feedback modulates the effect of varied TNFα stimulations on the NF-κB dynamics. We find that the different A20 feedback implementations exert similar but also model-specific effects. To demonstrate how the predicted distinct dynamic responses can be employed for model selection we compare our simulations results for incremental alterations of TNFα stimulation strength to corresponding experiments in Hela cells.

MATERIALS AND METHODS

Model Structures

In order to compare the three distinct implementations of the inhibitory mechanism of A20, we modularly designed three models. These models comprise an identical core module to which different upstream modules are attached (**Figures 1A,B**). The upstream modules are those proposed by Lipniacki et al. (2004), Ashall et al. (2009), and Murakawa et al. (2015) capturing different A20 feedback implementations. The overall models are hereafter referred to as model A–C.

The common core module of models A–C (**Figure 1B**) describes the reversible binding of free NF-κB and IκBα (reaction 1). Activated IKK (IKKactive) induces the IκBα degradation releasing NF-κB from the complex (reaction 5). Unbound NF-κB induces the transcription of IκBα mRNA (reaction 11), which is translated to IκBα (reactions 9). IκBα mRNA and IκBα protein degrade via reactions 7 and 4, respectively. In addition to IκBα mRNA, NF-κB induces the transcription of A20 mRNA (reaction 10). A20 mRNA is translated to A20 (reaction 8). A20 mRNA and protein are degraded via reactions 6 and 3, respectively. Taken together, the core module consists of five ordinary differential equations (ODEs) and one conservation relation for NF-κB. A detailed description of the corresponding rates and a list of the parameters are provided in the **Supplementary Information**.

The upstream module of model A (**Figure 1B**, left) comprises a very condensed representation of the activation of the IKK complex. The abundance of IKKactive increases in a TNFα-dependent and independent manner (reactions 13 and 14, respectively), both of which are inhibited by A20. IKKactive is inactivated via reaction 15.

In the upstream module of model B (**Figure 1B**, middle), IKK cycles between three distinct states: IKKneutral, IKKactive, and IKKinactive. TNFα stimulation converts IKKneutral into IKKactive (reaction 16), IKKactive is converted to IKKinactive (reaction 17) and IKKinactive is finally turned over to IKKneutral again (reaction 18). A20 inhibits this last reaction in a stimulus-sensitive manner.

The upstream module of model C (**Figure 1B**, right) includes the same states of IKK as described in model B, but IKKneutral, IKKactive, and IKKinactive do not interconvert in a cycle, i.e., obey a conservation relation. Instead, IKKneutral is continuously produced (reaction 24) and all three forms of IKK are subject to degradation (reactions 25–27). Similar to model B, TNFα stimulation in model C also converts IKKneutral into IKKactive (reaction 21), which in turn forms IKKinactive (reaction 23). In contrast to model B, model C includes an additional mechanism to convert IKKactive into IKKinactive (reaction 22). TNFα stimulation as well as A20 enhance this conversion.

Taken together, model A consist of one ODE in its upstream module in addition to the five ODEs and one conservation relation of NF-κB in the core module; model B incorporates two additional ODEs and an additional conservation relation of IKK in the upstream module; and model C includes three additional ODEs in its upstream module. Detailed descriptions of all three models are given in the **Supplementary Information**.

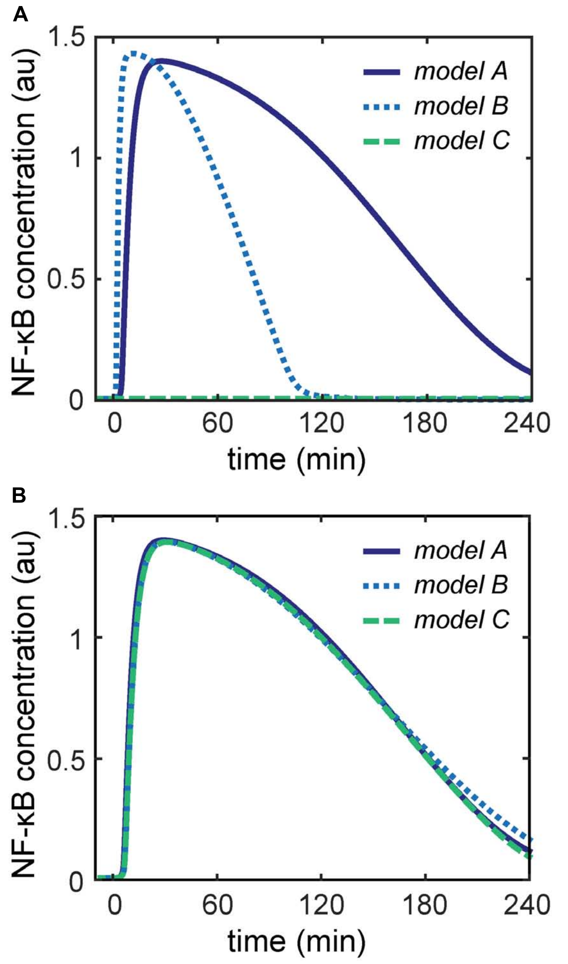


FIGURE 2 | NF-κB dynamics of the three models comprising the core module and the indicated upstream module. **(A)** Differences in NF-κB dynamics can be observed for the three models using the originally published parameters. **(B)** Nearly identical NF-κB dynamics can be observed for the three models with newly estimated parameters for the upstream modules.

Model Parameterizations

To parameterize the ODEs of the core module, we decided to use the parameters from our previously published model (Murakawa et al., 2015). This approach was based on two arguments. First, this model is based on a comprehensive data set characterizing the modulation of A20 feedback strength and its impact on NF-κB dynamics. Secondly, the core processes of this model perfectly match the reactions of the core module of our models A–C.

To parameterize the three different upstream modules of models A–C, we initially used the parameters published for the corresponding models (Lipniacki et al., 2004; Ashall et al., 2009; Murakawa et al., 2015). However, simulations of models A–C showed very diverse dynamics of unbound NF-κB in response to identical TNFα stimulation conditions (**Figure 2A**). For instance, the concentration of free NF-κB transiently increases in models A and B, but on a slower time scale in model A. In contrast, unbound NF-κB hardly increases upon TNFα stimulation in model C.

In order to compare models A–C directly, it is necessary that NF-κB exhibits the same dynamics upon TNFα stimulation in all three models. Thus, we estimated new parameters of the reactions in the upstream modules such that all components of the core module show the same dynamics in all three models. We used the D2D Toolbox (Raue et al., 2013) to estimate these parameters while keeping the parameters of the core module fixed. With this restriction on the parameters of the core module, we were able to reasonably minimize the parameter search space and obtain identical dynamics of the components of the core module. The details of the parameter estimation are explained in the **Supplementary Information**. Simulations of models A–C with these estimated parameters showed nearly identical dynamics of NF-κB activation upon TNFα stimulation (**Figure 2B**) and all remaining components of the core module (**Supplementary Figures S1, S2**).

Next, we checked whether the new parameterization changed the inhibitory effect of A20 on the activation of IKK. To do so, we simulated A20 knockout conditions by setting the A20 transcription rate k10 to zero and compared the resulting dynamics to those of wild-type conditions, i.e., using the reference value of k10 (**Supplementary Table S1**). The simulations show that the A20 knockout causes a prolonged increase in NF-κB, IKK and IκBα mRNA upon TNFα stimulation compared to wild-type (Lee et al., 2000) in all three models (**Supplementary Figures S3–S5**). The simulations furthermore show that the absence of A20 leads to a decrease in IκBα concentration in all three models. These results demonstrate that the parameterizations of the models A–C do represent the inhibitory effect of A20 on the activation of IKK.

Taken together, models A, B, and C were derived by modular design from an identical core module and different upstream modules specifying distinct implementations of the A20 feedback and TNFα stimulation. The models exhibit almost identical dynamics of their common model components, and show similar dynamical behavior in A20 knockout simulations.

Quantitative Characterization of the NF-κB Dynamics

To quantitatively compare the dynamics of unbound NF-κB between the models A–C, we used three established quantitative measures for signaling characteristics, in particular: (i) the maximal NF-κB concentration (x_{max}), (ii) the time of the maximal NF-κB concentration (t_{max}), and (iii) the response time (t_r) (**Figure 3**). The response time has been defined in Llorens et al. (1999), and quantifies the time required for a complete NF-κB response after stimulation. The function f is transformed to the gray line by taking the absolute gradient of f. The area above the transformed function is calculated and normalized by the steady state f* of the transformed function. While x_{max} and t_{max} describe the initial response of NF-κB to TNFα stimulation, t_r represents a normalized duration of NF-κB signaling and can therefore be used as a measure for the long-term dynamics.

Numerical Simulations

The model equations are listed in the **Supplementary Information**. Calculations were done with MathWorks Matlab R2013b. Steady state solutions were numerically obtained. Starting from those steady state solutions, the models are always simulated for 57,600 min in order to definitely reach a steady state and thus ensure convergence of the response time.

Experimental Methods

HeLa cells were stimulated with 10, 25, or 100 ng/ml TNFα (human recombinant TNFα, Alexis Corporation) for the time periods indicated (120, 100, 80, 60, 40, 20, and 10 min) or were left untreated. Following stimulation, cells were lysed in 20 mM Hepes pH = 7.9, 450 mM NaCl, 1 mM MgCl2, 0.5 mM EDTA pH = 8.0, 0.1 mM EGTA, 1% NP-40, 20% glycerol, supplemented with complete protease inhibitor mixture and Phosphostop (Roche Applied Science), 50 nM Calyculin A, 10 mM NaF, 10 mM β-glycerophosphate, 0.3 mM Na3VO4, and 1 mM Dithiothreitol. Lysates were centrifuged at 14,000 rpm for 10 min.

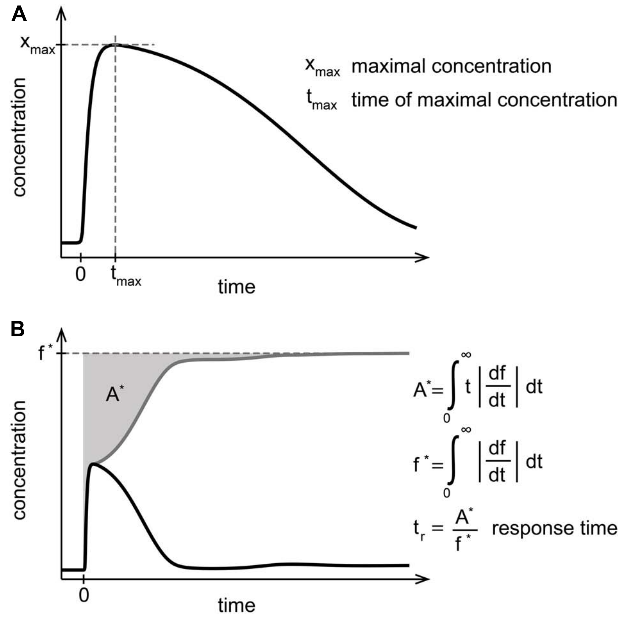


FIGURE 3 | Measures to quantify NF-κB dynamics. **(A)** The maximal concentration of NF-κB (x_{max}) and the time of the maximal concentration of NF-κB (t_{max}) characterize the initial NF-κB response. **(B)** The response time (t_r) defined in Llorens et al. (1999) is determined by the gray area (A*) normalized to the steady state (f*) of the absolute gradient of the dynamics of NF-κB. The response time quantifies the time required for the activation and deactivation of NF-κB upon stimulation and can be interpreted as a characterization of the NF-κB long-term behavior.

NF-κB DNA-binding activity was assayed by Electrophoretic Mobility Shift Assay (EMSA) as previously described (Stilmann et al., 2009).

EMSA quantification was made using the phosphor-imager Typhoon FLA 9500, GE Healthcare. Data were quantified using ImageQuant software. After background subtraction, the NF-κB band was normalized to a respective constant non-specific band.

RESULTS

Effects of Different A20 Feedback Strength on NF-κB Dynamics

As a starting point, we studied the impact of the A20 feedback on the NF-κB dynamics upon a constant TNFα stimulation. To do so, we varied the A20 feedback strength and studied its effects on the temporal change of the concentration of unbound NF-κB (hereafter denoted NF-κB) in all models. The strength of the A20 feedback is varied by multiplying the transcription rate constant of the A20 mRNA (k10) with a factor, i.e., feedback strength. A low value of the feedback strength corresponds to a weak negative feedback, whereas a high feedback strength results in a strong negative feedback. Local sensitivity analyses showed that a variation of the translation rate constants of A20 (k8) and of the transcription rate constant have a comparable effect on the three measures of the NF-κB dynamics (**Supplementary Figures S6–S8**). Thus, our choice to vary the transcription rate constant by a factor, i.e., the feedback strength, rather than the translation rate constant does not affect our conclusions.

The NF-κB dynamics of the models A–C for the A20 feedback strength 0.1 and 10 are shown in **Figure 4A**. In case of a high A20 feedback strength of factor 10, models B and C show a fast and transient increase of NF-κB concentration upon a constant TNFα stimulation (**Figure 4A** – top). In model A, NF-κB increases later and to a lesser extent compared to model B and C, yet it decreases to a similar final concentration. In the case of a low A20 feedback strength of factor 0.1 (**Figure 4A** – bottom), all three models show an almost identical increase in the NF-κB concentration. However, NF-κB decreases faster and to a lower final concentration in model C compared to model A and B. Comparing the simulations of the high with the low A20 feedback strength, all three models show a faster decrease in NF-κB in the case of high compared with low A20 feedback strength.

These results reflect the strong influence of the A20 feedback on the deactivation of NF-κB. A high A20 feedback strength causes a stronger and faster deactivation in all three models. Moreover, in model A a strong A20 feedback strength notably reduces and also delays NF-κB activation.

The IκBα Feedback Modulates the Effect of the A20 Feedback on NF-κB

Besides A20, IκBα is an important negative regulator of NF-κB dynamics. We next analyzed whether the interplay of these two feedbacks in the regulation of NF-κB dynamics is similar in the three models. To address this question, we varied the IκBα feedback strength in addition to that of A20. Similar to the A20 feedback strength, we multiplied the transcription rate constant of the IκBα mRNA (k11) by a factor to change the IκBα feedback strength.

The NF-κB dynamics of the three models for four exemplary combinations of different A20 and IκBα feedback strength are shown in **Figure 4B** (cases I–IV). The simulations show a rapid increase of NF-κB concentration upon TNFα stimulation for all models and in all four cases (I–IV), with one exception (model A, case I). The subsequent decrease of NF-κB concentration differs in strength and pace. For a combination of a high A20 feedback strength and a low IκBα feedback strength (case I), NF-κB concentrations in models B and C decrease to the half-maximum level at around 250 min whereas model A shows no NF-κB response to TNFα stimulation. When A20 and IκBα feedback strength are both low (case II), NF-κB concentration decreases at a much slower pace and to lesser extent than in case I for models B and C; here (case II) model A also shows a transient NF-κB activation. If the feedback strength of A20 and IκBα are high (case III), a fast increase can be observed that is followed by a nearly complete decrease of NF-κB concentration at 100 min for all models. For combinations of a high IκBα feedback strength with a low A20 feedback strength (case IV), the decrease in NF-κB concentration is slightly prolonged compared to case III, depending also on the model. These results are in agreement with our earlier finding that higher A20 feedback strength cause a faster and stronger decrease in NF-κB than lower A20 feedback strength (**Figure 4A**).

In the comparison of case I and case III, which both comprise the same A20 feedback strength but differ in their IκBα feedback strength, a stronger as well as faster decrease in the NF-κB concentration can be observed for high IκBα feedback strength. The comparison of case II and case IV yields a similar result, showing that a higher IκBα feedback strength leads to a faster and stronger decrease in NF-κB concentrations and therefore influencing its short-term and long-term dynamics.

In summary, both feedbacks lead to the deactivation of NF-κB after a transient increase. Thus, if only one of the two feedbacks is strong, it can compensate for the other. If A20 and IκBα feedback strength are both strong, the effect on the deactivation of NF-κB is enhanced resulting in an even faster and stronger NF-κB deactivation.

Beside these general observations, we find model-specific effects of the feedbacks. Most obviously, the maximal NF-κB activation and the deactivation pace seem to vary between the models. An interesting combination is a strong A20 with a low IκBα feedback strength (case I) for model A, which prevents an NF-κB response to TNFα stimulation.

Quantification of the Influences of the A20 and the IκBα Feedback on NF-κB Dynamics

To determine to what extent the models A–C differ in their NF-κB response under the various feedback strength, we quantified the dynamics of NF-κB by three measures: the maximal concentration of NF-κB, the time of the maximal concentration, and the response time (**Figure 3**). The first two measures characterize the initial NF-κB dynamics whereas the last measure characterizes the long-term NF-κB dynamics. For each model we then continuously varied the A20 and the IκBα feedback strength over a broad range of four orders of magnitude, covering very low (e.g., 0.01) as well as very high (e.g., 100) feedback strength (**Figure 4C**).

In model A, the maximal NF-κB concentration barely changes at A20 feedback strength below 1 (**Figure 4C** – first column, first row). In those cases, only an increase in the IκBα feedback strength leads to a decrease in the maximal concentration of NF-κB. For strong A20 feedback strength above 1, the A20 feedback can prevent the NF-κB response almost completely for a wide range of different IκBα feedback strength (**Figure 4C** – first row, black area). This is in agreement with case I in **Figure 4B** showing no NF-κB response for high A20 and low IκBα feedback strength. For A20 feedback strength below 1 in combination with a wide range of different IκBα feedback strength, the maximal concentration of NF-κB is reached in the first 80 min (**Figure 4C** – first column, second row – blue area). For A20 feedback strength above 1, an increase in the A20 feedback strength can lead to a delay in the time of the maximal concentration of NF-κB. Very high A20 feedback

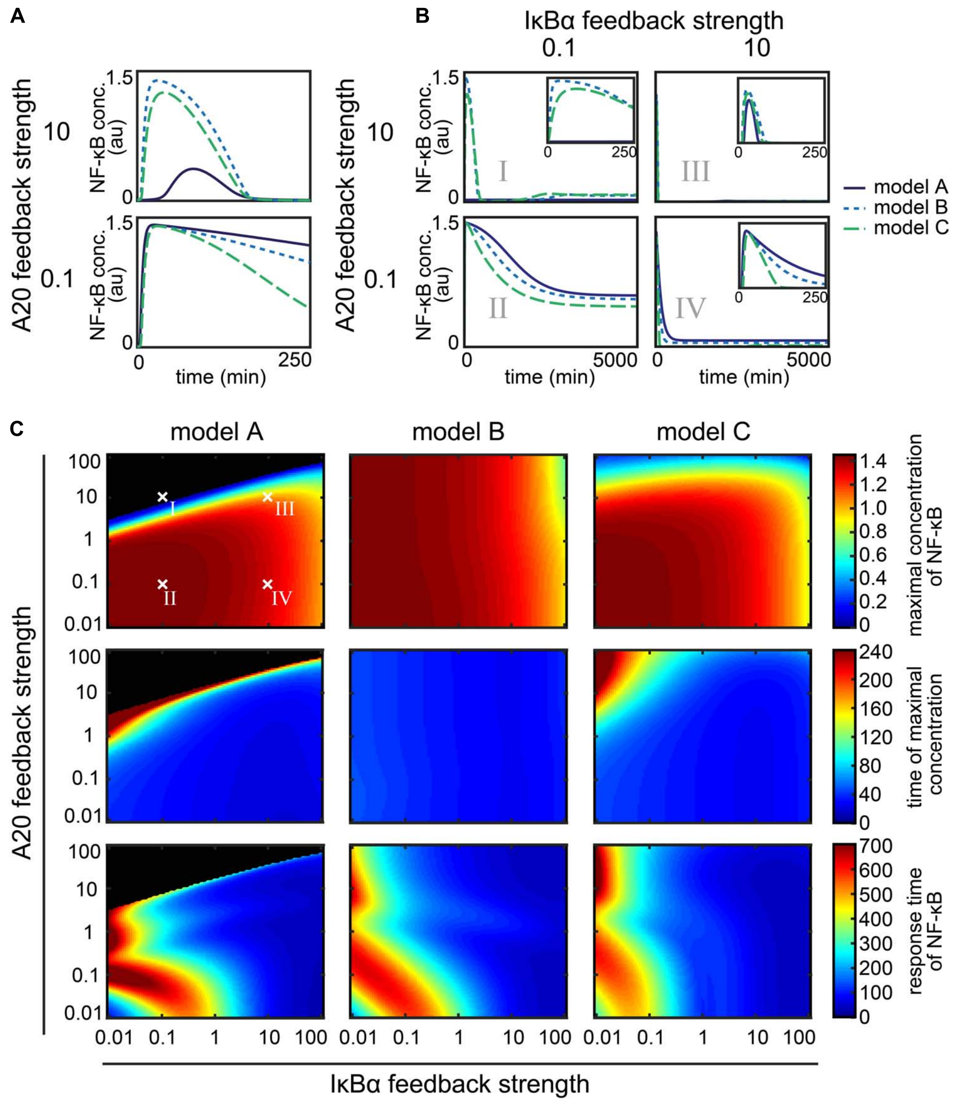


FIGURE 4 | Influence of the A20 feedback strength and the IκBα feedback strength on NF-κB dynamics. **(A)** NF-κB dynamics of the three models for two different A20 feedback strength. **(B)** NF-κB dynamics of the three models for four exemplary combinations of A20 and IκBα feedback strength. Insets zoom into the early time points of the dynamics. **(C)** The effect of the different combinations of feedback strength on the maximal concentration of NF-κB (first row), the time of the maximal concentration (second row), and the response time of NF-κB (third row) in the case of model A (first column), model B (second column) and model C (third column). The four exemplary combinations of feedback strength shown in panel B (I, II, III, and IV) are indicated. Black areas mark the combinations of feedback strength where hardly any NF-κB response is observed, i.e., the difference between maximal concentration of NF-κB and initial concentration of NF-κB is less than the threshold value of 0.001 μM.

strength completely diminish the NF-κB response. The effect of the A20 feedback on the response time of NF-κB is also modulated by the IκBα feedback (**Figure 4C** – first column, third row). The increase in the response time of NF-κB for confined combinations of low A20 and IκBα feedback strength is due to a prolonged higher concentration of NF-κB at later time points. The response time of NF-κB remains low for a wide range of different A20 feedback strength for IκBα feedback strength above 1. To summarize, the effects of the two feedbacks, A20 and IκBα, in model A can be subdivided into three main areas. The first area comprises combinations of A20 and IκBα feedback strength below 1. Those combinations result in a rapid but prolonged first peak of NF-κB and a higher NF-κB concentration at later time points similar to case II in **Figure 4B**. The second area is determined by high A20 feedback strength, where the NF-κB response is completely inhibited for low IκBα feedback strength similar to case I in **Figure 4B**. However, if the IκBα feedback strength is high, NF-κB remains responsive. The third area comprises high IκBα feedback strength resulting in a slightly decreased first peak of NF-κB and no response at later time points similar to case III and IV in **Figure 4B**.

In model B, the A20 feedback strength hardly influences the height and time of the maximal concentration of NF-κB. Both measures are mainly determined by the IκBα feedback strength (**Figure 4C** – second column, first and second row). However, the A20 feedback strength influences the response time of NF-κB (**Figure 4C** – second column, third row). Especially, if the A20 and IκBα feedback strength are both low, the NF-κB response time is higher. Thus, in model B the initial NF-κB response is mainly determined by the IκBα feedback, whereas the combination of both feedbacks influences the NF-κB dynamics at later time points.

In model C, an increase in the A20 feedback strength reduces the maximal concentration of NF-κB for A20 feedback strength above 1 (**Figure 4C** – third column, first row). For feedback strength below 1, the A20 feedback barely influences the maximal concentration of NF-κB. In those cases, an increase in the IκBα feedback strength can gradually decrease the maximal concentration of NF-κB. The time of the maximal concentration of NF-κB appears to be mainly robust toward changes in the two feedback strength (**Figure 4C** – third column, second row). Only combinations of A20 feedback strength above 1 and IκBα feedback strength below 0.1 delay the time of the maximal concentration of NF-κB. Considering the response time of NF-κB, the influence of the A20 feedback can be strongly modulated by the IκBα feedback (**Figure 4C** – third column, third row). The NF-κB response time remains low for IκBα feedback strength above 1 independent of the A20 feedback strength. For an IκBα feedback strength below 1, the A20 feedback strength can increase the NF-κB response time for A20 feedback strength either above 10 or for feedback strength between 1 and 0.1. To summarize, the effects of the two feedbacks in model C can be subdivided into three areas. The first area comprises combinations of A20 and IκBα feedback strength below 1. Those combinations result in a rapid, but prolonged first peak of NF-κB and a higher NF-κB concentration at later time points similar to case II in **Figure 4B**. The second area is confined by A20 feedback strength above 10 and IκBα feedback strength below 0.1 resulting in a reduced as well as a delayed maximal NF-κB concentration similar to case I in **Figure 4B**. The third area comprises IκBα feedback strength above 1 leading to a fast but decreased first peak of maximal NF-κB and no response at later time points similar to case III and IV in **Figure 4B**.

Altogether, the models show similar, but also different influences of the feedbacks on the NF-κB dynamics. For model A and C, the two negative feedbacks, IκBα and A20, have an impact on the initial dynamics. Both can independently reduce the maximal NF-κB concentration. However, in both models the two feedbacks are not completely redundant but have distinct functions in modulating the NF-κB response. If both feedback strength are below 1, the inhibitory effect of A20 and IκBα is weak. In that case, the initial NF-κB response is slightly delayed and a prolonged activation of NF-κB can be observed at later time points. If A20 feedback strength are high, the NF-κB response is completely inhibited in model A. In model C, a reduced as well as delayed NF-κB response can be observed. If the IκBα feedback strength is high, both models show a reduced but fast initial NF-κB increase and no response at later time points. To summarize, in models A and C both feedbacks inhibit the maximal concentration of NF-κB, but the A20 feedback delays the initial response and prolongs the response at later time points, whereas the IκBα feedback results in a faster initial activation and rapid deactivation of NF-κB. In contrast, in model B the initial NF-κB response is hardly influenced by the A20 feedback but mainly regulated by the IκBα feedback. Also in model B both feedbacks have an effect on the later phase of the NF-κB dynamics.

Characterization of the Interplay of TNFα Stimulation and A20 Feedback Strength

In all three considered mechanisms, the A20 feedback modulates the signal transduction of the TNFα stimulus toward the activation of IKK. We are therefore interested in the influence of the A20 feedback strength on the NF-κB response upon different strength of TNFα stimulation. To address this question, we simultaneously varied the stimulation strength of TNFα and the strength of the A20 feedback and quantified their influence on the maximal concentration of NF-κB, time of the maximal concentration and the response time of NF-κB (**Figure 5**). Here, the IκBα feedback strength is fixed to the value of 1.

In model A, variations in TNFα stimulation change the initial and long term dynamics of NF-κB (**Figure 5** – first column). In particular, an increase in TNFα stimulation strength leads to a faster and stronger increase in the maximal NF-κB value (**Figure 5** – first column, first and second row). This effect can be strongly modulated by the A20 feedback: for feedback strength above 1 a reduction and delay of the maximal NF-κB concentration can be observed. High A20 feedback strength above 10 result in a complete prevention of the NF-κB response for various TNFα stimulation strength (**Figure 5** – first column, black area). The response time of NF-κB is influenced by TNFα stimulation and A20 feedback strength in a complex way (**Figure 5** – first column, third row). For instance, for the combination of A20 feedback strength below 1 and TNFα stimulation strength above 1 the response time of NF-κB increases, indicating a prolonged NF-κB activation. In contrast, the combination of A20 feedback strength around 0.01 and TNFα stimulation strength above 10 leads to a decrease in the response time of NF-κB. The underlying reason is the change in the deactivation of NF-κB. For A20 feedback strength of 0.01 and TNFα stimulation strength of 100, NF-κB is not deactivated. Thus, NF-κB concentration does not decrease after its initial increase, resulting in a low response time (**Supplementary Figure S9**). However, for A20 feedback strength of 0.1 and TNFα stimulation strength of 100, NF-κB concentration slowly decreases after its initial increase, resulting in a high response time (**Supplementary Figure S9**).

In model B, the amount and time of the maximal concentration of NF-κB depend on the TNFα stimulation strength, but are mostly robust toward changes in A20 feedback strength (**Figure 5** – second column, first and second row). However, both TNFα stimulation strength and A20 feedback strength affect the response time of NF-κB (**Figure 5** – second column, third row). The effect is non-monotonous: low TNFα

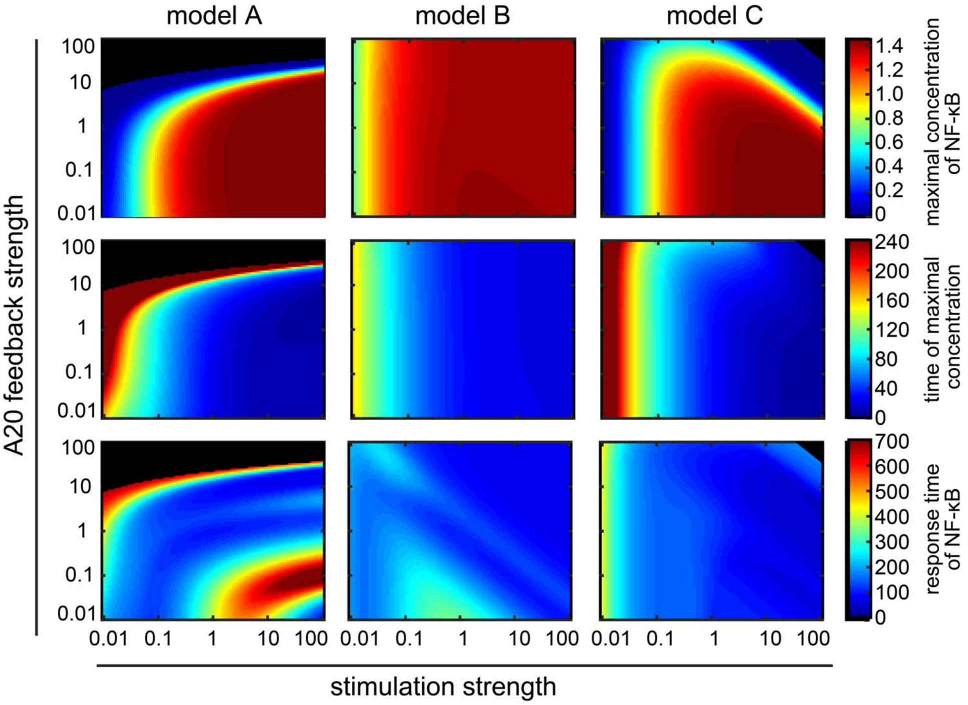


FIGURE 5 | Influence of A20 feedback strength and TNFα stimulation strength on NF-κB dynamics. NF-κB dynamics of model A (first column), model B (second column) and model C (third column) are characterized by the maximal concentration of NF-κB (first row), the time of the maximal concentration of NF-κB (second row) and the response time of NF-κB (third row). Black areas mark combinations of A20 feedback strength and TNFα stimulation strength with hardly any observable NF-κB response; the difference between maximal and initial NF-κB concentrations is less than 0.001 μM.

stimulation strength between 0.1 and 1 and very low A20 feedback strength below 0.1 show an increase in the response time of NF-κB, indicating a prolonged activation of NF-κB. However, in the case of TNFα stimulation strength between 10 and 100, a decrease in the response time is observed.

In model C, the maximal concentration of NF-κB and the timing of its peak mostly depend on TNFα stimulation strength (**Figure 5** – third column, first and second row). A20 feedback strength can lead to a reduction and a slight delay of the maximal NF-κB concentration for high TNFα stimulation strength. In particular, if A20 feedback strength as well as TNFα stimulation strength are high, the maximal concentration of NF-κB decreases and can result in a complete prevention of the NF-κB response (**Figure 5** – third column, black area). The response time of NF-κB mainly depends on TNFα stimulation strength and hardly on A20 feedback strength (**Figure 5** – third column, third row).

In conclusion, the maximal NF-κB concentration and its timing, are strongly determined by the TNFα stimulation strength in all models. In models A and C the A20 feedback can strongly modify that impact. However, in model B, we see no significant effect of the A20 feedback on the amount and time of maximal NF-κB. The effect of the TNFα stimulation strength and the A20 feedback on the long-term dynamics is more complex. However, if we consider the effect of TNFα stimulation (for factors > 1) and a given A20 feedback strength (factor = 1), we observe opposite effects in the models: while a higher TNFα stimulation strength leads to an increase of the response time in model A, such a stimulus increase would cause a decrease in the response time in models B and C.

Comparison of Simulations With Experimental Data for the Effect of Varied TNFα Stimulation Strength

The qualitative differences between the models suggest an experimental setup to scrutinize the A20 feedback implementations. To predict the outcome of such an experiment, we simulated the NF-κB dynamics of the models A–C in response to three different TNFα concentrations (**Figure 6A**). We selected TNFα stimulation because changes in TNFα concentration are easier to perform experimentally than changes in A20 feedback strength. Our simulations predict for model A that NF-κB levels remain high for stimulation with 100 ng/ml TNFα compared with 10 ng/ml TNFα at later time points (**Figure 6A**). In contrast, in models B and C, NF-κB levels decrease faster at later time points upon stimulation with 100 ng/ml TNFα compared to 10 ng/ml TNFα. These predictions are independent of the assumed A20 feedback strength (**Supplementary Figure S10**) and are furthermore verified by simulations of the models published by Lipniacki et al. (2004), Ashall et al. (2009), and Murakawa et al. (2015) (**Supplementary Figure S11**).

We compared our model predictions to experimental data applying 10, 25, and 100 ng/ml TNFα to HeLa cells. The time course measurements of NF-κB's DNA-binding activity by EMSA

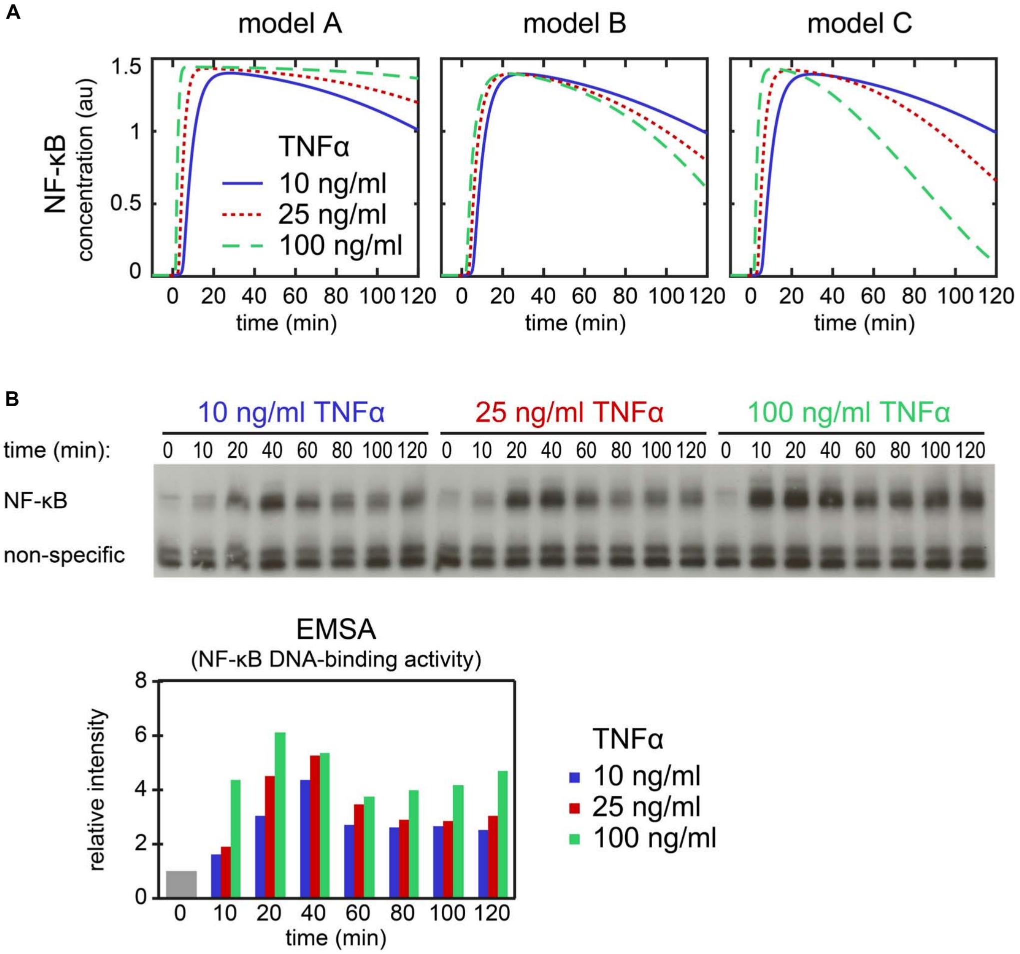


FIGURE 6 | Dynamics of NF-κB upon stimulation with different TNFα concentrations. **(A)** Simulation of NF-κB assuming a stimulation with 10 ng/ml (solid line), 25 ng/ml (dotted line), and 100 ng/ml TNFα (dashed line) in model A (left), model B (middle) and model C (right). **(B)** Exemplary EMSA experiment measuring NF-κB DNA-binding activity over a time course of 120 min in HeLa cells upon stimulation with 10, 25, and 100 ng/ml TNFα. The histogram shows the quantification of the EMSA experiment. The mean value of the relative intensities at $t = 0$ is set to 1 and used as a normalization for all other values. Two replicate experiments are shown as **Supplementary Figure S12**.

showed NF-κB dynamics as predicted for model A but not model B or C (**Figure 6B**). The comparison of model results and experiments thus suggests that in HeLa cells the implementation of the A20 feedback structure of model A is appropriate to describe the effect of A20 on the dynamics of NF-κB.

DISCUSSION

In this study, we developed a modular modeling approach to analyze the impact of different A20 inhibition mechanisms on the dynamics of NF-κB. In particular, we compared three distinct implementations of the A20 feedback by combining upstream modules of available models with a common core pathway module. By fitting the resulting models to a published comprehensive experimental data set, we derive models with quantitatively comparable NF-κB dynamics. When analysing the effect of variations of the strength of the A20 and IκBα feedbacks, as well as of TNFα stimulation in these models, we observe similarities, but also model-specific differences. Increasing IκBα feedback strength attenuate the initial as well as the long-term NF-κB response in all three models, that is, reduce the maximum and response time, respectively. Increasing A20 feedback strength reduce the maximum and duration of the NF-κB response in models A and C. In model A, the NF-κB response is even completely diminished for very high A20 feedback strength. However, in model B the A20 feedback has no impact on the initial dynamics. Moreover, our simulations predicted that changes in the TNFα stimulation strength influence initial and long-term dynamics of NF-κB. Here, we observed qualitative differences in the long-term NF-κB response between the different models. We used these predictions for an experimental validation in HeLa cells. The experimental observations support model A, but not model B or C in this cell type.

Models A–C differ in the implementation of the A20 feedback. We compared the effect of this feedback implementation for a carefully derived parametrization of the models. While the detailed NF-κB dynamics might change for other model parametrizations, we expect the effect of the model structure to more generally valid. In all three models, A20 acts conjointly with the stimulus in order to inhibit IKK activation. Model A includes in addition a basal IKK activation rate that is inhibited by A20 (reaction 14). Such a composite, non-linear description of the inhibitory influence of A20 seems necessary to reproduce the NF-κB dynamics of HeLa cells. This indicates that the regulation of IKK activity by A20 in this cell type may result from a combination of several mechanisms and is thus more complex than anticipated. Indeed, A20 seems to fulfill multiple functions *in vivo*, such as a deubiquitinating activity mediated by its N-terminal ovarian tumor (OTU) domain and an E3 ubiquitin ligase activity mediated by its C-terminal zinc finger domain (Lork et al., 2017). These distinct functions of A20 may regulate the activity of upstream signal mediators and constitute potential mechanisms that may explain the complex non-linearity in the signal transduction from TNFα stimulation to IKK activation (Hymowitz and Wertz, 2010). In a stochastic pathway model the different A20 effects have been combined to better explain experimental data (Lipniacki et al., 2007). A recent analysis of temperature effects on the NF-κB pathway also highlights the importance of the A20 feedback and the necessity to extend and modify its implementation in model B (Ashall et al., 2009; Harper et al., 2018). Moreover, it will be interesting to explore the role of additional negative regulators on the pathway, e.g., the deubiquitinating enzymes CYLD and OTULIN (Lork et al., 2017) as well as the effect of the cross-talk with the non-canonical pathway (Ashall et al., 2009; Yilmaz et al., 2014; Mukherjee et al., 2017).

Our analyses of the three models revealed redundant but also distinct functions of the two negative feedbacks, A20 and IκBα. This confirms and extends earlier findings by Werner et al. (2008), demonstrating distinct roles of the two feedbacks in a very detailed pathway model. In that publication, IκBα has been reported to modulate mostly the initial NF-κB response while A20 mainly shapes the late response. In our current study, we characterize the output based on quantitative measures for a wide range of different feedback strength. We find that the IκBα feedback fine-tunes the initial NF-κB response in all models. However, it can also influence the response-time and therefore the long-term dynamics. The A20 feedback has different effects in models A, B, and C. In models A and C, it modulates the initial as well as long-term dynamics. Moreover, in model A it has a bimodal on-off effect on the NF-κB response, i.e., preventing the NF-κB response at high A20 feedback strength. While our analysis revealed a lower sensitivity of model B to changes in the A20 feedback, a comprehensive analysis (**Supplementary Figures S6–S8**) showed comparable sensitivities of all three models to parameter changes in general, only the distribution of the sensitivities between processes differs in the models.

The non-redundant functions of the two negative feedbacks could be due to their structural properties: the two feedbacks are interlocked, with the IκBα feedback serving as an inner feedback loop and the A20 feedback as an outer feedback loop. Previous studies indicted distinct functions of interlocked feedback loops with respect to the oscillatory behavior of a system (Nguyen, 2012; Baum et al., 2016). Here, a weak or strong outer feedback loop may cause an on or off response, respectively, independent of the strength of the inner feedback loop. However, the inner feedback loop can fine-tune the response in the case of a weak outer feedback loop. Such interlocked feedback loops are very common regulatory motifs in signaling pathways in general (Batchelor et al., 2011; Benary et al., 2015; Kochanczyk et al., 2017; Zhang et al., 2017).

Taken together, our quantitative modular modeling approach employs the regulation of NF-κB signaling by the A20 feedback as an example case to study the impact of different implementations of an inhibition mechanism on the model's response to perturbations. Comparing the simulations of the three models A–C to experimental data suggests that model A is an appropriate choice to describe TNFα stimulation in HeLa cells. Our results emphasize the need to further explore the molecular details of processes upstream of IKK regulation.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JM and JW contributed to the conceptualization and design of study. JM contributed to the development, simulation and analysis of ODE models, design and implementation of computer code. JM, UB, and JW contributed to the data interpretation and curation. II and SÇ contributed to the experimental work. CS contributed to the supervision of experimental work. JW contributed to the supervision of project. JM, UB, CS, and JW contributed to the preparation of manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The project was supported by a grant from the German Federal Ministry of Education and Research BMBF (Project ProSiTu, 0316047A) awarded to JW and CS and by the Personalized Medicine Initiative "iMed" of the Helmholtz Association to JW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at https://www.biorxiv.org (Mothes et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00896/full#supplementary-material

REFERENCES

Ashall, L., Horton, C. A., Nelson, D. E., Paszek, P., Harper, C. V., Sillitoe, K., et al. (2009). Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. *Science* 324, 242–246. doi: 10.1126/science.1164860

Basak, S., Behar, M., and Hoffmann, A. (2012). Lessons from mathematically modeling the NF-kappaB pathway. *Immunol. Rev.* 246, 221–238. doi: 10.1111/j.1600-065x.2011.01092.x

Batchelor, E., Loewer, A., Mock, C., and Lahav, G. (2011). Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* 7:488. doi: 10.1038/msb.2011.20

Baum, K., Politi, A. Z., Kofahl, B., Steuer, R., and Wolf, J. (2016). Feedback, mass conservation and reaction kinetics impact the robustness of cellular oscillations. *PLoS Comput Biol.* 12:e1005298. doi: 10.1371/journal.pcbi.1005298

Benary, U., and Wolf, J. (2019). Controlling nuclear NF-kappaB dynamics by beta-TrCP-insights from a computational model. *Biomedicines* 7:40. doi: 10.3390/biomedicines7020040

Benary, U., Kofahl, B., Hecht, A., and Wolf, J. (2015). Mathematical modelling suggests a differential impact of beta-transducin repeat-containing protein paralogues on Wnt/beta-catenin signalling dynamics. *FEBS J.* 282, 1080–1096. doi: 10.1111/febs.13204

Cheong, R., Hoffmann, A., and Levchenko, A. (2008). Understanding NF-kappaB signaling via mathematical modeling. *Mol. Syst. Biol.* 4:192. doi: 10.1038/msb.2008.30

De, A., Dainichi, T., Rathinam, C. V., and Ghosh, S. (2014). The deubiquitinase activity of A20 is dispensable for NF-kappaB signaling. *EMBO Rep.* 15, 775–783. doi: 10.15252/embr.201338305

Fagerlund, R., Behar, M., Fortmann, K. T., Lin, Y. E., Vargas, J. D., and Hoffmann, A. (2015). Anatomy of a negative feedback loop: the case of IkappaBalpha. *J. R. Soc. Interf.* 12:0262.

Harper, C. V., Woodcock, D. J., Lam, C., Garcia-Albornoz, M., Adamson, A., Ashall, L., et al. (2018). Temperature regulates NF-kappaB dynamics and function through timing of A20 transcription. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5243–E5249.

Hayden, M. S., and Ghosh, S. (2012). NF-kappaB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* 26, 203–234. doi: 10.1101/gad.183434.111

Hinz, M., and Scheidereit, C. (2014). The IκB kinase complex in NF−κB regulation and beyond. *EMBO Rep.* 15, 46–61. doi: 10.1002/embr.201337983

Hoffmann, A., Levchenko, A., Scott, M. L., and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* 298, 1241–1245. doi: 10.1126/science.1071914

Huxford, T., Huang, D. B., Malek, S., and Ghosh, G. (1998). The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. *Cell* 95, 759–770.

Hymowitz, S. G., and Wertz, I. E. (2010). A20: from ubiquitin editing to tumour suppression. *Nat. Rev. Cancer* 10, 332–341. doi: 10.1038/nrc2775

Kochanczyk, M., Kocieniewski, P., Kozlowska, E., Jaruszewicz-Blonska, J., Sparta, B., Pargett, M., et al. (2017). Relaxation oscillations and hierarchy of feedbacks in MAPK signaling. *Sci. Rep.* 7:38244.

Lee, E. G., Boone, D. L., Chai, S., Libby, S. L., Chien, M., Lodolce, J. P., et al. (2000). Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. *Science* 289, 2350–2354. doi: 10.1126/science.289.5488.2350

Lipniacki, T., and Kimmel, M. (2007). Deterministic and stochastic models of NFkappaB pathway. *Cardiovasc. Toxicol.* 7, 215–234. doi: 10.1007/s12012-007-9003-x

Lipniacki, T., Paszek, P., Brasier, A. R., Luxon, B., and Kimmel, M. (2004). Mathematical model of NF-kappaB regulatory module. *J. Theor. Biol.* 228, 195–215. doi: 10.1016/j.jtbi.2004.01.001

Lipniacki, T., Puszynski, K., Paszek, P., Brasier, A. R., and Kimmel, M. (2007). Single TNFalpha trimers mediating NF-kappaB activation: stochastic robustness of NF-kappaB signaling. *BMC Bioinform.* 8:376. doi: 10.1186/1471-2105-8-376

Llorens, M., Nuno, J. C., Rodriguez, Y., Melendez-Hevia, E., and Montero, F. (1999). Generalization of the theory of transition times in metabolic pathways: a geometrical approach. *Biophys. J.* 77, 23–36. doi: 10.1016/s0006-3495(99)76869-4

Longo, D. M., Selimkhanov, J., Kearns, J. D., Hasty, J., Hoffmann, A., and Tsimring, L. S. (2013). Dual delayed feedback provides sensitivity and robustness to the NF-kappaB signaling module. *PLoS Comput. Biol.* 9:e1003112. doi: 10.1371/journal.pcbi.1003112

Lork, M., Verhelst, K., and Beyaert, R. (2017). CYLD, A20 and OTULIN deubiquitinases in NF-kappaB signaling and cell death: so similar, yet so different. *Cell Death Differ.* 24, 1172–1183. doi: 10.1038/cdd.2017.46

Mothes, J., Busse, D., Kofahl, B., and Wolf, J. (2015). Sources of dynamic variability in NF-kappaB signal transduction: a mechanistic model. *Bioessays* 37, 452–462. doi: 10.1002/bies.201400113

Mothes, J., Ipenberg, I., Arslan, S. Ç, Benary, U., Scheidereit, C., and Wolf, J. (2019). A quantitative modular modeling approach reveals the consequences of different A20 feedback implementations for the NF-kB signaling dynamics. *bioRxiv* [Preprint].

Mukherjee, T., Chatterjee, B., Dhar, A., Bais, S. S., Chawla, M., Roy, P., et al. (2017). A TNF-p100 pathway subverts noncanonical NF-kappaB signaling in inflamed secondary lymphoid organs. *EMBO J.* 36, 3501–3516. doi: 10.15252/embj.201796919

Murakawa, Y., Hinz, M., Mothes, J., Schuetz, A., Uhl, M., Wyler, E., et al. (2015). RC3H1 post-transcriptionally regulates A20 mRNA and modulates the activity of the IKK/NF-kappaB pathway. *Nat. Commun.* 6:7367.

Nguyen, L. K. (2012). Regulation of oscillation dynamics in biochemical systems with dual negative feedback loops. *J. R. Soc. Interf.* 9, 1998–2010. doi: 10.1098/rsif.2012.0028

Perkins, N. D. (2012). The diverse and complex roles of NF-kappaB subunits in cancer. *Nat. Rev. Cancer* 12, 121–132. doi: 10.1038/nrc3204

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelker, M., Kaschek, D., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One* 8:e74335. doi: 10.1371/journal.pone.0074335

Skaug, B., Chen, J., Du, F., He, J., Ma, A., and Chen, Z. J. (2011). Direct, noncatalytic mechanism of IKK inhibition by A20. *Mol. Cell* 44, 559–571. doi: 10.1016/j.molcel.2011.09.015

Stilmann, M., Hinz, M., Arslan, S. C., Zimmer, A., Schreiber, V., and Scheidereit, C. (2009). A nuclear poly(ADP-ribose)-dependent signalosome confers DNA damage-induced IkappaB kinase activation. *Mol. Cell* 36, 365–378. doi: 10.1016/j.molcel.2009.09.032

Werner, S. L., Kearns, J. D., Zadorozhnaya, V., Lynch, C., O'Dea, E., Boldin, M. P., et al. (2008). Encoding NF-kappaB temporal control in response to TNF: distinct roles for the negative regulators IkappaBalpha and A20. *Genes Dev.* 22, 2093–2101. doi: 10.1101/gad.1680708

Wertz, I. E., Newton, K., Seshasayee, D., Kusam, S., Lam, C., Zhang, J., et al. (2015). Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. *Nature* 528, 370–375. doi: 10.1038/nature16165

Williams, R., Timmis, J., and Qwarnstrom, E. (2014). Computational models of the NF-KB signalling pathway. *Computation* 2:131. doi: 10.3390/computation2040131

Yilmaz, Z. B., Kofahl, B., Beaudette, P., Baum, K., Ipenberg, I., Weih, F., et al. (2014). Quantitative dissection and modeling of the NF-kappaB p100-p105 module reveals interdependent precursor proteolysis. *Cell Rep.* 9, 1756–1769. doi: 10.1016/j.celrep.2014.11.014

Zambrano, S., Bianchi, M. E., and Agresti, A. (2014). A simple model of NF-kappaB dynamics reproduces experimental observations. *J. Theor. Biol.* 347, 44–53.

Zhang, Z. B., Wang, Q. Y., Ke, Y. X., Liu, S. Y., Ju, J. Q., Lim, W. A., et al. (2017). Design of tunable oscillatory dynamics in a synthetic NF-kappaB signaling circuit. *Cell Syst.* 5, 460-70.e5.

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 28 July 2020
doi: 10.3389/fphys.2020.00803

Development and Analysis of a Quantitative Mathematical Model of Bistability in the Cross Repression System Between APT and SLBO Within the JAK/STAT Signaling Pathway

Alyssa Berez[1], *Bradford E. Peercy*[1*] *and Michelle Starz-Gaiano*[2]

[1] *Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD, United States,*
[2] *Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, MD, United States*

Edited by:
Denis Tsygankov,
Georgia Institute of Technology,
United States

Reviewed by:
Didier Gonze,
Université libre de Bruxelles, Belgium
Nathan Weinstein,
Universidad Nacional Autnoma de
Mexico, Mexico

***Correspondence:**
Bradford E. Peercy
bpeercy@umbc.edu

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: *31 March 2020*
Accepted: *17 June 2020*
Published: *28 July 2020*

Citation:
Berez A, Peercy BE and Starz-Gaiano M (2020) Development and Analysis of a Quantitative Mathematical Model of Bistability in the Cross Repression System Between APT and SLBO Within the JAK/STAT Signaling Pathway. Front. Physiol. 11:803. doi: 10.3389/fphys.2020.00803

Cell migration is a key component in development, homeostasis, immune function, and pathology. It is important to understand the molecular activity that allows some cells to migrate. *Drosophila melanogaster* is a useful model system because its genes are largely conserved with humans and it is straightforward to study biologically. The well-conserved transcriptional regulator Signal Transducer and Activator of Transcription (STAT) promotes cell migration, but its signaling is modulated by downstream targets Apontic (APT) and Slow Border Cells (SLBO). Inhibition of STAT activity by APT and cross-repression of APT and SLBO determines whether an epithelial cell in the *Drosophila* egg chamber becomes motile or remains stationary. Through mathematical modeling and analysis, we examine how the interaction of STAT, APT, and SLBO creates bistability in the Janus Kinase (JAK)/STAT signaling pathway. In this paper, we update and analyze earlier models to represent mechanistically the processes of the JAK/STAT pathway. We utilize parameter, bifurcation, and phase portrait analyses, and make reductions to the system to produce a minimal three-variable quantitative model. We analyze the manifold between migratory and stationary steady states in this minimal model and show that when the initial conditions of our model are near this manifold, cell migration can be delayed.

Keywords: JAK/STAT, *Drosophila melanogaster*, border cell migration, mathematical model, bistability

1. INTRODUCTION

The acquisition of cellular migration plays a critical role in both normal and pathological development. A better understanding of the processes cells undergo as they transition from a stationary state to a migratory state is thus of broad interest. Studying the mechanics of how cohorts of cells move together introduces additional complexities, and existing models of collectively migrating cells differ greatly (Peercy and Starz-Gaiano, 2020; Stuelten et al., 2018; Aman and Piotrowski, 2010; Saadin and Starz-Gaiano, 2016; Olson and Nechiporuk, 2018; Leonard and Taneyhill, 2020; Friedl and Mayor, 2017). To study cell migration, some scientists turn to an experimental model system amenable to both genetic analysis and live imaging. In *Drosophila* *melanogaster* during oogenesis, a set of cells called border cells develop within a layer of follicle cells and later become migratory, leaving the nearby epithelial cells behind as they move to the oocyte (Montell et al., 2012; Saadin and Starz-Gaiano, 2016). Experimentalists have discovered much of the molecular regulation that governs this process and what causes some border cells to become motile while others remain stationary, including the primary biochemical and molecular pathways. We are interested in advancing a mathematical model of these pathways, which could have implications on acquisition of cell motility in animals in general.

The Janus Kinase/Signal Transducer and Activator of Transcription (JAK/STAT) signaling pathway has been shown by previous studies to be crucial in the motility of border cells, as well as in stem cells and immune response (Montell et al., 2012; Arbouzova and Zeidler, 2006; Trivedi and Starz-Gaiano, 2018; Amoyel and Bach, 2012; Amoyel et al., 2014). The JAK/STAT pathway is well-conserved from fruit flies to humans. Anterior polar cells in the *Drosophila* egg chamber (see **Figures 1**, **2A**) secrete the cytokine Unpaired (UPD), which acts as the ligand for the transmembrane Domeless receptor in neighboring follicle cells. UPD is predicted to form a gradient across the adjacent cells (Van De Bor et al., 2011; Xi et al., 2003; Starz-Gaiano et al., 2008). The binding of UPD to Domeless activates JAK, leading to the phosphorylation of the activated JAK/receptor complex. The activated complex then recruits and phosphorylates STAT. The phosphorylated STAT dimerizes, moves to the nucleus, and acts as a transcription factor for specific target genes.

A key target of STAT is the C/EBP transcription factor Slow Border Cells (Slbo) (Montell et al., 1992; Starz-Gaiano et al., 2008). Cells specified by high levels of SLBO to become the border cells respond to chemoattractants that activate two receptor tyrosine kinases (RTKs), which turn on signaling cascades that promote directional movement (Montell et al., 2012; Stuelten et al., 2018; Saadin and Starz-Gaiano, 2016; Friedl and Mayor, 2017). Downstream of the RTKs, guided motility is governed in border cells largely through regulation of the actinomyosin cytoskeleton via the Rho GTPase, Rac, and myosin phosphorylation states, with coordinated changes in cell-cell adhesion mediated by E-cadherin (Stuelten et al., 2018; Saadin and Starz-Gaiano, 2016; Montell et al., 2012; Chen et al., 2020). The follicle cells receiving sufficiently high levels of UPD—not necessarily just those in closer proximity to the polar cells depending on the extracellular geometry (Manning et al., 2015)—turn on a higher level of STAT activity and become motile border cells while nearby cells with lower levels do not. Interestingly, high STAT activity in ovarian follicle cell is sufficient to induce motility in usually stationary lateral follicle cells (Silver and Montell, 2001).

We focus on the protein products from two genes activated by STAT: Apontic (APT) and SLBO. SLBO promotes migratory behavior and an insufficient amount of SLBO prevents motility (Montell et al., 1992). APT protein is a transcription factor that downregulates the function of JAK/STAT and SLBO, and thus inhibits migration (Starz-Gaiano et al., 2008). APT acts as a feedback inhibitor on the JAK/STAT pathway, and this process is mediated by APT's activation of a microRNA that reduces STAT protein and activity (Yoon et al., 2011). APT also activates expression of *Socs36E*, which downregulates STAT signaling via a degradation pathway (Monahan and Starz-Gaiano, 2015). APT and SLBO exhibit cross-repressional behavior (Starz-Gaiano et al., 2008, 2009). APT directly represses *slbo* transcription while SLBO only decreases the level of expression of APT protein. The dominating protein in a given cell determines the cell fate: stationary or motile. This creates what appears to be a bistable system, as cells that receive intermediate amounts of STAT have the potential to reach either fate (Starz-Gaiano et al., 2009; Rorth, 1994). While it is reasonable based on the non-linearities in the system to model this cell fate regulation as bistability, *in vivo* experiments to support this are challenging and we are not aware of any particular experiments that have been done to confirm bistability. By identifying the conditions under which bistability occurs, modeling can help to design experimental protocols to confirm bistability.

We base our model on the mechanistic model built by Ge et al. (2012). Focusing on the cross-repression system of APT and SLBO, they built a mathematical model using elementary interactions to identify which components of the system are sufficient for bistability. Depending on the strength of the UPD signal and thus the level of STAT activity, each border cell can become motile (SLBO dominates) or remain stationary (APT dominates). Cells with an intermediate level of STAT activity may fall above or below the threshold necessary for mobility.

Ge and Stonko created a 15-variable model including many mechanisms of STAT regulation (see **Figure 2A**) as well as the cross-repression system of APT and SLBO (see **Figure 2B**) with sufficient elements, specifically cooperativity in SLBO repressing *apt* mRNA translation, to cause bistability. We do not know the mechanism for this, but we suspect SLBO activates expression of a microRNA that mediates the effect. Ge et al. (2012) identified several miRNAs in the *Drosophila* genome that have upstream binding sites for SLBO activation and seed sequences that would target the apt mRNA 3′ untranslated regions.

To describe briefly the full model, in Equations (1) and (2) JAK (J) is altered to an enzymatically active state, J^*, in the presence of UPD (U). Further in Equation (1) J^* enzymatically activates two STAT monomers (S) to create the activated STAT dimer (S_2^*) shown for STAT variables and complexes in Equations (3)–(6) with c_1 denoting the complex between activated JAK (J^*) and STAT (S). Equation (5) represents APT sequestering S_2^* in c_2. In Equations (7) and (8) APT (A) and SLBO (B) are produced from their mRNA and degraded, while A binding with S_2^* is accounted for. The mRNA production of *stat* (m_σ), *apt* (m_α), and *slbo* (m_β), is shown in Equations (7)–(9) based on a constant basal production and transcribing or probabilistic fraction based on non-transcribing *stat* (σ), *apt* (α), and *slbo* (β) respectively, along with degradation. In Equation (9) B cooperatively enhances m_α degradation, while in Equations (10) and (11) A enhances m_σ and m_β degradation. The gene state dynamics are shown in Equations (12)–(15), with S_2^* inducing transcription of σ, α, and β, while binding with A puts *slbo* into a repressed state (β^R). Model variables for Equations (1)–(15) from Ge et al. (2012) are in **Table 1**.

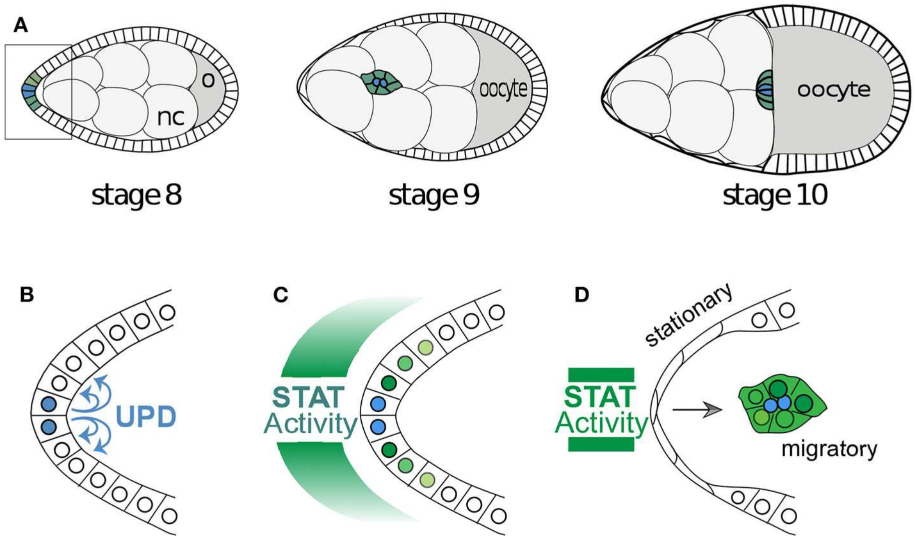


FIGURE 1 | Egg development and the migration of border cells. **(A)** A cartoon of the development of a *Drosophila* egg chamber and the movement of the border cells between the nurse cells (nc). The outer edge of the egg chamber is made up of epithelial follicle cells and the box outlines the cells in **(B)**. **(B)** The signaling molecule Unpaired is secreted from the polar cells and **(C)** induces a gradient of STAT activity across the anterior epithelium. Often the follicle cells very close to the polar cells assume the identity of border cells and **(D)** become motile and migrate as a cluster toward the oocyte. The cells with below-threshold levels of STAT activity shut the signaling off in a switch like manner.

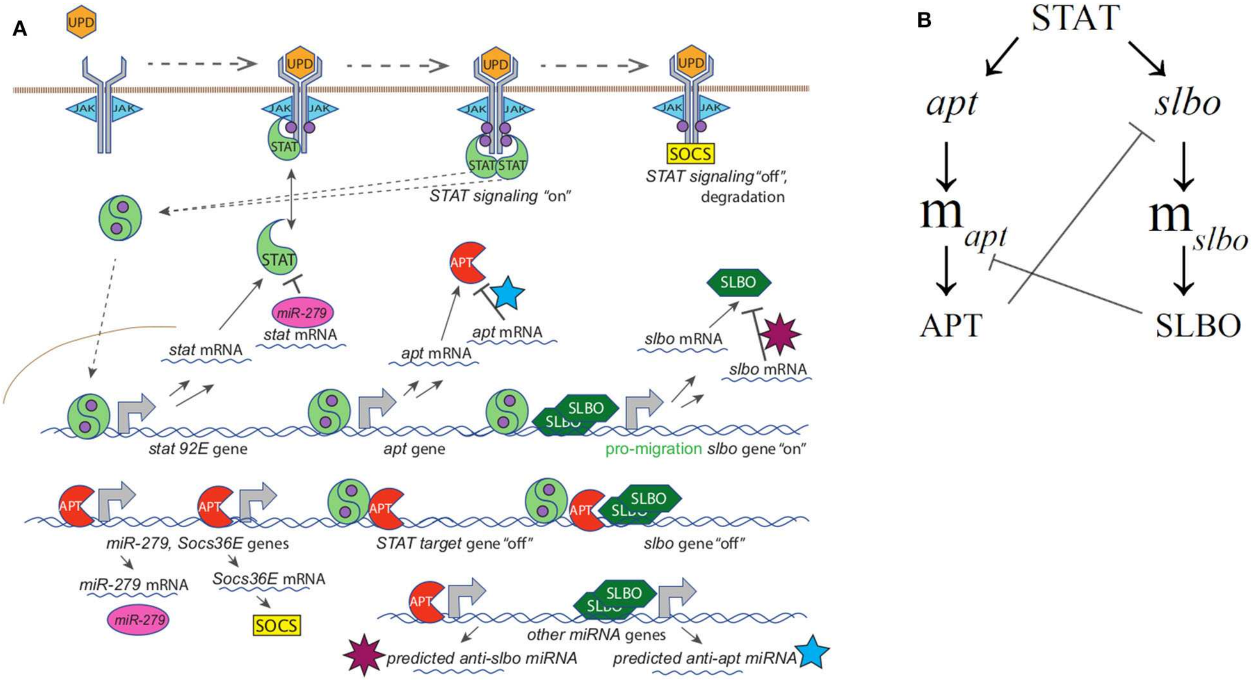


FIGURE 2 | (A) Full mechanistic diagram of JAK/STAT pathway **(B)** Cross-repression system of APT and SLBO. STAT activates *apt* and *slbo* transcription leading to the production of *apt* and *slbo* mRNA and translation into APT and SLBO proteins. SLBO represses *apt* translation while APT suppresses *slbo* transcription and function. High APT levels make the cell stationary and high SLBO levels make the cell motile. However, which state dominates cannot be determined from this qualitative diagram alone.

TABLE 1 | 15-variable model variables.

Variable	Description and Units
J^*	Activated JAK protein (nM)
J	JAK protein (nM)
S	STAT protein monomer (nM)
c_1	J^* and S complex (nM)
c_2	S_2^* and A complex (nM)
S_2^*	Activated STAT protein dimer (nM)
A	APT protein (nM)
B	SLBO protein (nM)
m_α	APT mRNA (nM)
m_β	SLBO mRNA (nM)
m_σ	STAT mRNA (nM)
α	Proportion of inactive apt genes
β	Proportion of inactive slbo genes
β^R	Proportion of repressed slbo genes
σ	Proportion of inactive stat genes

The 15-variable model is:

$$
\begin{aligned}
\frac{dJ^*}{dt} &= k^f_{UJ}UJ - k^b_{UJ}J^* - k^f_{c_1}J^*S^2 + k^b_{c_1}c_1 + k_{c_1}c_1 && (1)\\
\frac{dJ}{dt} &= -k^f_{UJ}UJ + k^b_{UJ}J^* && (2)\\
\frac{dS}{dt} &= -2k^f_{c_1}J^*S^2 + 2k^b_{c_1}c_1 + 2k_{S_2^*}S_2^* + k_s m_\sigma - \delta_S S && (3)\\
\frac{dc_1}{dt} &= k^f_{c_1}J^*S^2 - k^b_{c_1}c_1 - k_{c_1}c_1 && (4)\\
\frac{dc_2}{dt} &= k^f_{S_2^*A}S_2^*A - k^b_{S_2^*A}c_2 && (5)\\
\frac{dS_2^*}{dt} &= k_{c_1}c_1 - k_{S_2^*}S_2^* - k^f_{S_2^*A}S_2^*A + k^b_{S_2^*A}c_2 && (6)\\
\frac{dA}{dt} &= k_A m_\alpha - \delta_A A - k^f_{S_2^*A}S_2^*A + k^b_{S_2^*A}c_2 && (7)\\
\frac{dB}{dt} &= k_B m_\beta - \delta_B B && (8)\\
\frac{dm_\alpha}{dt} &= k_{m_\alpha}(1-\alpha) - \delta_{m_\alpha}m_\alpha + m^o_\alpha - \delta_{B_\alpha}B^2 m_\alpha && (9)\\
\frac{dm_\beta}{dt} &= k_{m_\beta}(1-\beta-\beta^R) - \delta_{m_\beta}m_\beta + m^o_\beta - \delta_{A_\beta}Am_\beta && (10)\\
\frac{dm_\sigma}{dt} &= k_{m_\sigma}(1-\sigma) - \delta_{m_\sigma}m_\sigma + m^o_\sigma - \delta_{A_\sigma}Am_\sigma && (11)\\
\frac{d\alpha}{dt} &= -k^f_\alpha S_2^*\alpha + k^b_\alpha(1-\alpha) && (12)\\
\frac{d\beta}{dt} &= -k^f_\beta S_2^*\beta + k^b_\beta(1-\beta-\beta^R) && (13)\\
\frac{d\beta^R}{dt} &= k^f_{\beta^R}A\beta - k^b_{\beta^R}\beta^R && (14)\\
\frac{d\sigma}{dt} &= -k^f_\sigma S_2^*\sigma + k^b_\sigma(1-\sigma) && (15)
\end{aligned}
$$

In this paper we analyze and adapt the Ge and Stonko model so that a minimal reduction retains the dynamics of the detailed model. In section 2 we describe the methods we use to establish parameters and for bifurcation analysis. In section 3 we develop the minimal reduced model. In section 4 we analyze bifurcation results, identify the critical stable manifold separating the migratory and stationary steady states, validate experimental results, and compare our model to previous models. We apply this model in the interesting case of controlling microRNA-mediated degradation of *stat* mRNA via APT and show that delays in STAT activation, even to the point of activation failure within a biophysical time span, are possible due to the proximity of the critical UPD level to a limit point bifurcation. We conclude with possible experimentation that could test and improve our model.

2. MATERIALS AND METHODS

In this section, we describe the methods used to analyze our minimal reduced model developed below. We use bifurcation and parameter analysis and identify the stable manifold between the stationary and motile steady states.

2.1. Time Course and Bifurcation Analysis

We use XPPAUT (Ermentrout, 2002) to create time course simulations and bifurcation diagrams and to analyze the bistability of the reduced model. We use the stiff integration method to solve our system of ODEs. A bifurcation occurs when a small change in parameter values results in a qualitative change in a system. Bifurcation analysis allows us to identify limit points treating UPD as a parameter, and helps to identify the separatrix between our steady states.

2.2. Establishing Parameters

In order to establish a more biophysically realistic model for this study, we researched existing literature to find data to establish parameter values. For some parameters we were able to find data specific to the JAK/STAT pathway or *Drosophila*. For other parameters, related pathways were used to obtain data relevant to this model.

We were able to identify published values for general translation and transcription rates and applied the established rates to the lengths of JAK (encoded by *hopscotch*), STAT (encoded by *Stat92E*), APT, and SLBO genes and proteins (Hargrove et al., 1991; Lewin, 2004). The lengths are found in the fly genome database, Flybase (Thurmond et al., 2019). Protein and mRNA degradation rates have a wide range of average values so JAK, STAT, APT, and SLBO are assumed to conform to this range (Guido et al., 2006; Harris et al., 2011; Nicholson and Nicola, 2013). The protein to DNA binding and dissociation rates of JAK and STAT have been identified in general but not specifically for APT and SLBO (Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Nicholson and Nicola, 2013).

Many signaling pathways, including the JAK/STAT pathway, are regulated by microRNAs (miRNAs) (Lui et al., 2015; Yoon et al., 2011). Ge et al. (2012) identified several microRNAs annotated in the genome (Thurmond et al., 2019; Betel et al., 2007) that have seed sequences corresponding to *apt* and *slbo*, suggesting they are also regulated by this mechanism. This regulation can occur through mRNA degradation, translational inhibition, or other means, making the parameters corresponding to miRNA kinetics difficult to assign. Ge and Stonko addressed this by condensing the various processes by which miRNA can affect mRNA into one degradation rate. We identified a parameter value for this combined effect through information from the model established in Yoon et al. (2011). The rate that *slbo* transitions in and out of its repressed state was also hard to identify due to lack of data, so we utilized the rate given in Ge and Stonko which was adapted from the rate for a different repressor in Harris et al. (2011) and information from Rorth (1994). The base levels of STAT and the total amount of JAK present in border cells were estimated from Yoon et al. (2011), Starz-Gaiano et al. (2008), and McGregor et al. (2002). Lastly, for the rates of STAT independent mRNA production, we again used the original parameter values from Starz-Gaiano et al. (2009). For *slbo*, this rate is most likely negligible. However, *apt* can be activated by means other than STAT. The protein Eyes Absent (EYA) can also activate transcription of *apt* (Starz-Gaiano et al., 2009).

2.3. Manifolds Separating Cell Fate Basins of Attraction

One goal in developing the three-variable model was to be able to fully understand the manifold that separates the steady states in the model. For a level of UPD in the bistable region, cells can either become motile or remain stationary depending on the initial conditions of the system.

We visually identified the manifolds by labeling initial conditions according to the steady state to which they converge. This allows us to see the basins of attraction for each steady state. These two stable steady states, one where SLBO dominates and the cell becomes motile and one where APT dominates and the cell remains stationary, are listed in the appendix with values for each variable. In our three-dimensional system we are able to approximate the manifold by fitting a surface to the boundary between the basins of attraction.

A graphical representation of the boundary manifold was achieved by creating a 3-d grid of initial conditions and determining to which steady state each converged within 500 min, a time deemed reasonable from experimental data. Any initial conditions that did not converge by this time were identified as lying near the manifold. We then used the MATLAB curve fitting toolbox to fit a surface to these points, creating an approximation of the manifold between the steady states.

3. DEVELOPING THE REDUCED MODEL

We began by reducing the fixed STAT cross-repression system of APT and SLBO (Equations, 7–10, 12–14) to a two-variable model. In the process of researching biophysically realistic parameters we discovered that the binding rate of STAT to target genes appears to be several orders of magnitude faster than any other process in the system, as seen in **Table 2** (Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010). For example, k^f_α, k^f_β, and $k^f_{\beta^R}$ are at least three orders of magnitude faster than the translation and transcription kinetics. Additionally, α, β, and β^R reach equilibrium significantly faster than the other variables. We used time-scale analysis to reduce the system. We made a quasi-steady state approximation for α, β, and β^R and set those derivatives equal to zero. This allowed us to solve Equations (12), (13), and (14) for $\alpha^* = 1 - \alpha$ and $\beta^* = 1 - \beta - \beta^R$ in terms of APT protein and STAT dimer:

$$\alpha^* = \frac{\frac{S_2^*}{K_\alpha}}{\frac{S_2^*}{K_\alpha} + 1} \tag{16}$$

$$\beta = \frac{1}{\frac{S_2^*}{K_\beta} + 1 + \frac{A}{K_{\beta^R}}} \tag{17}$$

$$\beta^R = \frac{k^f_{\beta^R}}{k^b_{\beta^R}} A\beta = \frac{\frac{A}{K_{\beta^R}}}{\frac{S_2^*}{K_\beta} + 1 + \frac{A}{K_{\beta^R}}} \tag{18}$$

$$\beta^* = 1 - \beta - \beta^R = \frac{\frac{S_2^*}{K_\beta}}{\frac{S_2^*}{K_\beta} + 1 + \frac{A}{K_{\beta^R}}} \tag{19}$$

$$\text{with } K_\alpha = \frac{k^b_\alpha}{k^f_\alpha}, \quad K_\beta = \frac{k^b_\beta}{k^f_\beta}, \quad K_{\beta^R} = \frac{k^b_{\beta^R}}{k^f_{\beta^R}}$$

Our parameter values indicate that the mRNA processes occur at least twice as fast as the protein processes. This makes a quasi-steady state approximation for m_α and m_β plausible. Thus, we have a two-variable model where only APT and SLBO are dynamic.

The two-variable model is:

$$\frac{dA}{dt} = k_A \frac{k_{m_\alpha}\alpha^* + m^o_\alpha}{\delta_{m_\alpha} + \delta_{B_\alpha} B^2} - \delta_A A \tag{20}$$

$$\frac{dB}{dt} = k_B \frac{k_{m_\beta}\beta^* + m^o_\beta}{\delta_{m_\beta} + \delta_{A_\beta} A} - \delta_B B \tag{21}$$

$$m_\alpha = \frac{k_{m_\alpha}\alpha^* + m^o_\alpha}{\delta_{m_\alpha} + \delta_{B_\alpha} B^2} \tag{22}$$

$$m_\beta = \frac{k_{m_\beta}\beta^* + m^o_\beta}{\delta_{m_\beta} + \delta_{A_\beta} A} \tag{23}$$

After establishing the cross-repressional system of APT and SLBO, we reintroduced STAT dynamics to the model. Now we reduce the STAT activation system (Equations 1–6, 11, 15) through a number of assumptions.

First, we ignored the theoretical APT-STAT complex (c_2) as its effects of APT sequestering STAT do not affect the steady state structure of the model, which we proved analytically. Then

TABLE 2 | Three-variable system parameter values.

Parameter	Symbol	Value	Units	Citation
Binding rate of UPD to JAK	k^f_{UJ}	0.0133	min^{-1}	Manning et al., 2015; Ghiglione et al., 2002; Wright et al., 2011; Ward et al., 1995; Hilton and Nicola, 1992
Dissociation rate of UPD to JAK	k^b_{UJ}	0.1	$nM \cdot min^{-1}$	Ghiglione et al., 2002; Wright et al., 2011; Ward et al., 1995; Hilton and Nicola, 1992
Binding rate of J^* complex to 2 STAT monomers	$k^f_{c_1}$	1	min^{-1}	Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010
Dissociation rate of J^* complex to 2 STAT monomers	$k^b_{c_1}$	0.1	$nM \cdot min^{-1}$	Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010
Rate of S^*_2 leaving J^*	k_{c_1}	100	$nM \cdot min^{-1}$	Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010
Dedimerization rate of S^*_2	$k_{S^*_2}$	0.1	$nM^{-1} \cdot min^{-1}$	Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010
Rate of STAT translation	k_S	3	min^{-1}	Hargrove et al., 1991; Lewin, 2004; Thurmond et al., 2019
Rate of degradation of STAT	δ_S	0.1	min^{-1}	Harris et al., 2011; Nicholson and Nicola, 2013
Rate of transcription of *stat*	k_{m_σ}	1	min^{-1}	Hargrove et al., 1991; Lewin, 2004; Thurmond et al., 2019
Rate of degradation of *stat* mRNA	δ_{m_σ}	0.2	min^{-1}	Guido et al., 2006; Harris et al., 2011
Base level of *stat* mRNA	m^o_σ	0.5	$nM \cdot min^{-1}$	Ge et al., 2012; Silver and Montell, 2001; Ghiglione et al., 2002
Degradation rate of *stat* mRNA due to miRNA induced by APT	$\delta_{A\sigma}$	0.05	$nM^{-1} \cdot min^{-1}$	Yoon et al., 2011
Binding rate of S^*_2 to *stat*	k^f_σ	1	min^{-1}	Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010
Dissociation rate of S^*_2 to *stat*	k^b_σ	2	$nM \cdot min^{-1}$	Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Nicholson and Nicola, 2013; Karsten et al., 2006; Ekas et al., 2010
Total amount of JAK	J_T	0.15	nM	Ge et al., 2012; McGregor et al., 2002
Rate of APT translation	k_A	0.298	min^{-1}	Hargrove et al., 1991; Lewin, 2004; Thurmond et al., 2019
Rate of transcription of *apt*	k_{m_α}	0.54	$nM \cdot min^{-1}$	Hargrove et al., 1991; Lewin, 2004; Thurmond et al., 2019
Rate of STAT independent production of *apt* mRNA	m^o_α	0.52	$nM \cdot min^{-1}$	Starz-Gaiano et al., 2009
Rate of degradation of APT	δ_A	0.04	min^{-1}	Starz-Gaiano et al., 2008; Ge et al., 2012; Harris et al., 2011
Rate of degradation of *apt* mRNA	δ_{m_α}	0.086	min^{-1}	Guido et al., 2006; Harris et al., 2011; Starz-Gaiano et al., 2008
Degradation rate of *slbo* mRNA due to miRNA induced by APT	δ_{A_β}	0.1	$nM^{-1} \cdot min^{-1}$	Ge et al., 2012; this work
Binding rate of S^*_2 to *apt*	k^f_α	100	min^{-1}	Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Starz-Gaiano et al., 2008
Dissociation rate of S^*_2 to *apt*	k^b_α	0.66	$nM \cdot min^{-1}$	Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002
Rate of SLBO translation	k_B	0.312	min^{-1}	Hargrove et al., 1991; Lewin, 2004; Thurmond et al., 2019
Rate of transcription of *slbo*	k_{m_β}	0.538	$nM \cdot min^{-1}$	Hargrove et al., 1991; Lewin, 2004; Thurmond et al., 2019
Rate of STAT independent production of *slbo* mRNA	m^o_β	0.03	$nM \cdot min^{-1}$	Starz-Gaiano et al., 2009
Rate of degradation of SLBO	δ_B	0.04	min^{-1}	Harris et al., 2011; Rorth, 1994; Rorth et al., 2000
Rate of degradation of *slbo* mRNA	δ_{m_β}	0.086	min^{-1}	Guido et al., 2006; Harris et al., 2011
Degradation rate of *apt* mRNA due to miRNA induced by SLBO	δ_{B_α}	0.5	$nM^{-2} \cdot min^{-1}$	Ge et al., 2012; this work
Binding rate of S^*_2 to *slbo*	k^f_β	100	min^{-1}	Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Starz-Gaiano et al., 2008
Dissociation rate of S^*_2 to *slbo*	k^b_β	0.66	$nM \cdot min^{-1}$	Halford and Marko, 2004; Parsaeian et al., 2013; Yang et al., 2002; Starz-Gaiano et al., 2008
Rate *slbo* transitions into repressed state	$k^f_{\beta^R}$	100	min^{-1}	Harris et al., 2011; Rorth, 1994
Rate *slbo* transitions out of repressed state	$k^b_{\beta^R}$	0.522	$nM \cdot min^{-1}$	Harris et al., 2011; Rorth, 1994

we used the Michaelis-Menten approximation for the activated JAK (J^*) conversion of two STAT molecules to an activated STAT dimer (S^*_2). This eliminates the JAK-STAT complex (c_1) and condenses the conversion. We assumed conservation of JAK to eliminate unactivated JAK (J) by defining a constant total JAK as $J_T = J + J^*$. We also assumed that UPD (U) activation of JAK and STAT activation are fast so J^* and S^*_2 can be solved for by quasi-steady state approximations. Lastly,

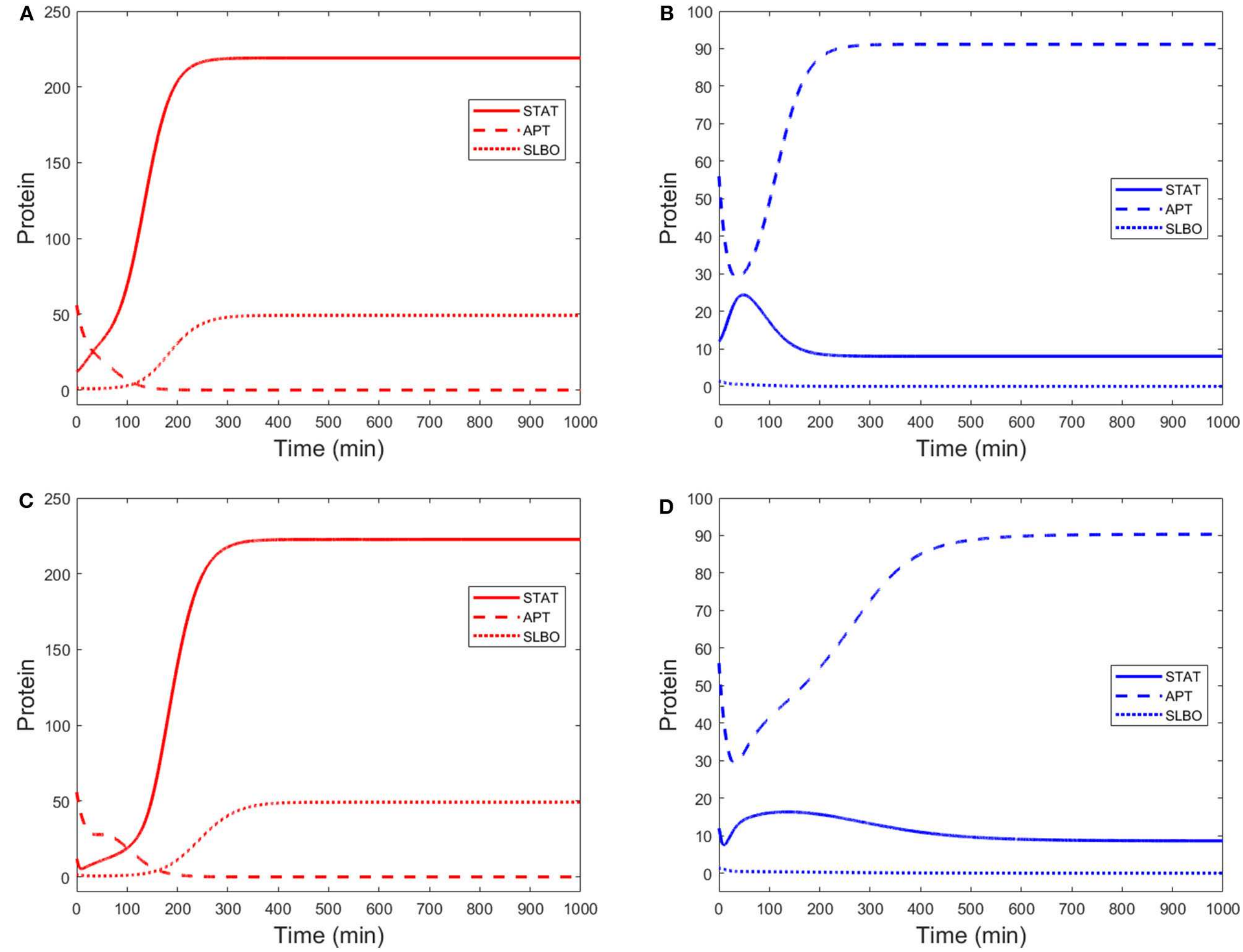


FIGURE 3 | Time courses of STAT, APT, and SLBO with initial conditions STAT = 12, APT = 56, SLBO = 1.5 in the 3-variable model for **(A)** motile steady state (UPD = 4) and **(B)** stationary steady state (UPD = 1) and the 15-variable model for **(C)** motile steady state (UPD = 4) and **(D)** stationary steady state (UPD = 1).

similar to the reduction made for *apt* and *slbo*, the inactive *stat* gene state (σ) and *stat* mRNA (m_σ) were solved by quasi-steady state approximations. These assumptions gave us the following additional approximations:

$$\sigma = \frac{1}{\frac{S_2^*}{K_\sigma}+1} \text{ with } K_\sigma = \frac{k_\sigma^b}{k_\sigma^f} \tag{24}$$

$$v_{max} = k_{c_1} J^* \tag{25}$$

$$k_m = \sqrt{\frac{k_{c_1}^b + k_{c_1}}{k_{c_1}^f}} \tag{26}$$

$$J^* = \frac{k_{UJ}^f U J_T}{k_{UJ}^b + k_{UJ}^f U} \tag{27}$$

$$S_2^* = \frac{1}{k_{S_2^*}} \frac{v_{max} S^2}{S^2 + k_m^2} \tag{28}$$

Thus, producing our minimal three-variable model in APT, SLBO, and STAT:

$$\frac{dA}{dt} = k_A \frac{k_{m_\alpha}\alpha^* + m_\alpha^o}{\delta_{m_\alpha} + \delta_{B_\alpha} B^2} - \delta_A A \tag{29}$$

$$\frac{dB}{dt} = k_B \frac{k_{m_\beta}\beta^* + m_\beta^o}{\delta_{m_\beta} + \delta_{A_\beta} A} - \delta_B B \tag{30}$$

$$\frac{dS}{dt} = k_S \frac{k_{m_\sigma}(1-\sigma) + m_\sigma^o}{\delta_{m_\sigma} + \delta_{A\sigma} A} - \delta_S S \tag{31}$$

4. RESULTS

4.1. Bifurcation Analysis

Time courses of STAT, APT, and SLBO show the difference between the motile and stationary steady states in **Figures 3A,B**. **Figures 3C,D** shows the comparable results for the original 15-variable model. The two models converge to slightly different steady state values. This is due to the methods used to reduce the

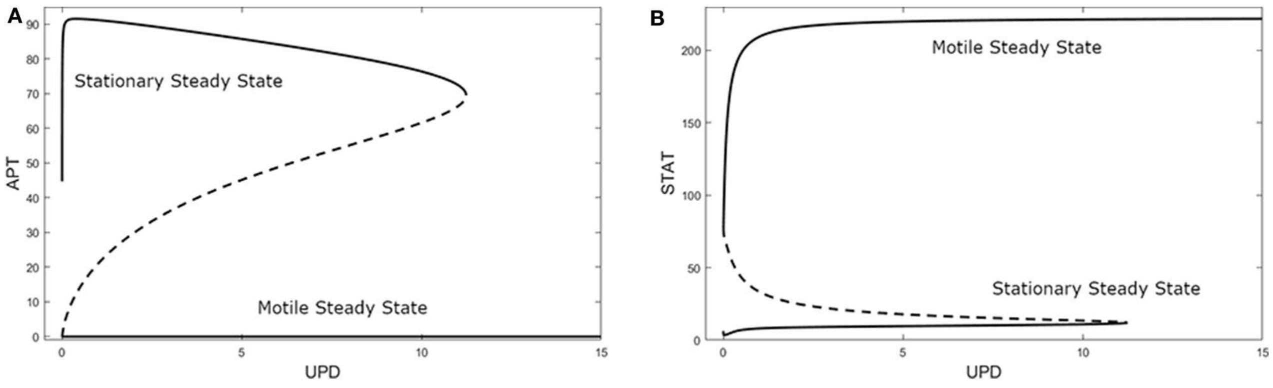


FIGURE 4 | Bifurcation diagrams of **(A)** APT and **(B)** STAT against UPD in the three-variable model. A solid line indicates a stable steady state, a dashed line indicates an unstable steady state. Limit point bifurcations (where the sold and dashed lines meet) create "knees" where the model jumps from one steady state to the other. At U = 0 APT dominates.

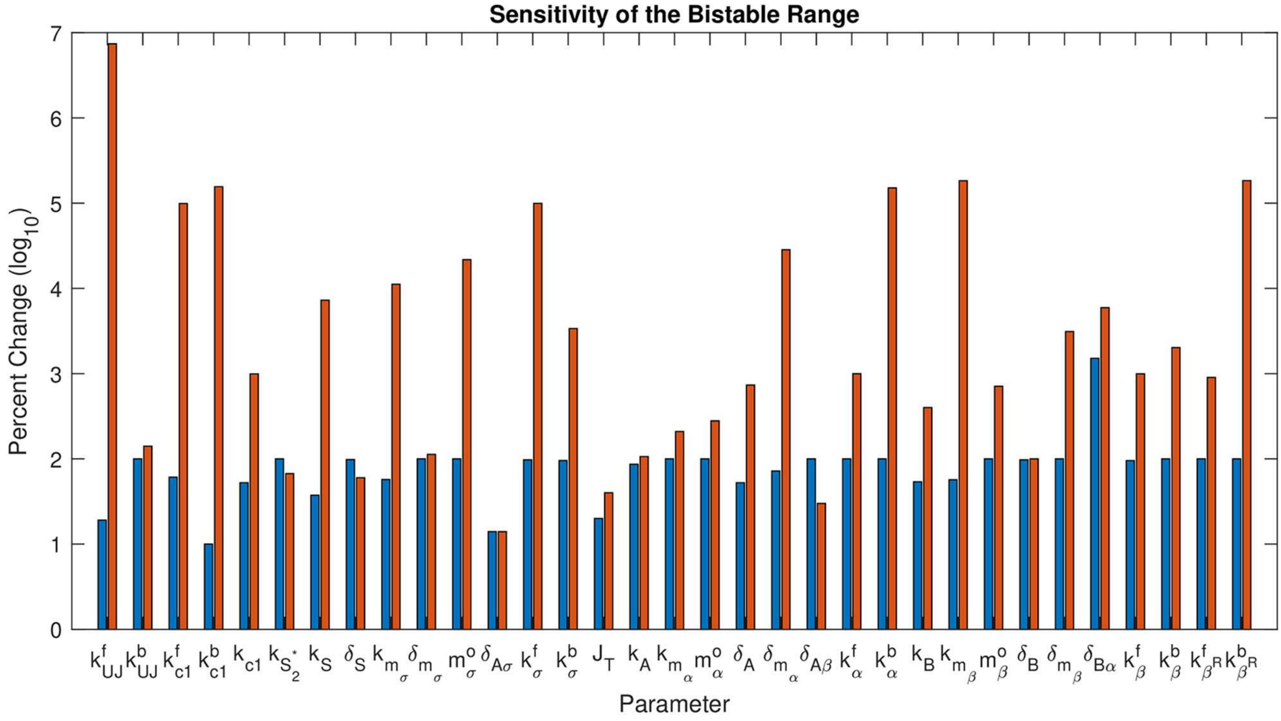


FIGURE 5 | Sensitivity of the Bistable Range. This shows $\log_{10}$ of percent change in each parameter that maintains bistability. The first bar (blue) for each parameter is the negative change, the second (red) is the positive change.

STAT dynamics in the 3-variable model. Exact steady state values can be found in the **Supplementary Materials**.

A bifurcation diagram of APT against UPD revealed a non-trivial state when $U = 0$ with a high level of APT, as seen in **Figure 4A**. This can be interpreted as the system being predisposed to the stationary cell fate until UPD and thus STAT is high enough, at UPD $>$ 11.24 nM. As UPD approaches 0 on the upper branch, APT stops at a value of 44.73 nM. Since APT is also positively affected by STAT via UPD, very low UPD will decrease the steady state levels of APT but not to zero, leaving the stationary state the only steady state at very low UPD. Here the system encounters a limit point bifurcation and with a higher UPD level would transition to low APT and a motile cell fate.

Additionally, with the same parameters as **Figure 4A**, the STAT bifurcation (**Figure 4B**) shows that if UPD begins at a high level in a cell, the cell will remain in the motile state even as UPD decreases to a very low level. This can be seen in experiments that dissociate polar cells and border cells. The border cells continue to migrate until their UPD level presumably drops below the threshold level, i.e., as they get too far away from the polar cells

(Starz-Gaiano et al., 2008; Cai et al., 2014). This threshold also exists in the opposite direction—as the level of UPD increases in an epithelial cell, it becomes a migratory border cell once the threshold is reached (Manning et al., 2015).

The bistability in the model depends on two mechanisms. We show that the predominate non-linearity is in SLBO cooperative repression of *apt* mRNA translation. We also show the repressed state for the *slbo* gene induced by APT contributes to bistability (Starz-Gaiano et al., 2008). Bifurcation diagrams for STAT, APT, and SLBO with each combination of these criteria are presented as **Figures S1–S3**. As quadratic non-linearity in SLBO is made linear and the ability to reach the *slbo* gene repressed state is eliminated, the stationary basin of attraction becomes smaller until bistability is lost.

4.2. Parameter Values

Throughout the research on parameter values, the goal was to develop a range of realistic parameters to test in our model. There are two reasons why a range of values is desirable. First, from the biological point of view, many biological processes do not occur at a constant rate. Second, heterogeneity in cellular parameters will likely lead to some parameter variety. A range of average values is thus both more appropriate and more consistent with experimental data. We were able to test the robustness of the dynamics over the ranges of parameters to see if the model behavior matches experimentally observed outcomes.

The parameter values established through research from a variety of experimental systems and testing for bistability are shown in **Table 2**. The range of values identified for each parameter demonstrates a robust region of bistability. This adds confidence in the decision to use some parameter values that were established from a range of possible values.

Figure 5 shows the percent change on a $\log_{10}$ scale in each parameter that maintains bistability. The exact range for each parameter is listed in **Table 3**. The most sensitive parameter is $\delta_{A\sigma}$. This makes sense biologically, as $\delta_{A\sigma}$ controls APT induced miRNA degradation of *stat*. Changing this rate directly affects whether a presumptive border cell produces enough STAT activity to pass the threshold needed for motility given a particular amount of UPD. APT has multiple levels of control on STAT, but this may be the most sensitive because there are multiple other regulators at the protein level. We explore the effects of this parameter on the outcome of the system below. The range of possible values for each parameter supports the robustness of the JAK/STAT pathway as a bistable system.

TABLE 3 | Bistable Range of each parameter.

Parameter	Bistable start	Bistable end
k_{UJ}^{f}	0.007	985
k_{UJ}^{b}	0	0.241
$k_{c_1}^{f}$	0.389	993
$k_{c_1}^{b}$	0	156
k_{c_1}	47.43	1093
$k_{S_2^*}$	0	0.167
k_S	1.875	221
δ_S	0.0013	0.16
k_{m_σ}	0.428	113
δ_{m_σ}	0	2.46
m_σ^o	0	109
$\delta_{A\sigma}$	0.043	0.057
k_σ^f	0.028	996
k_σ^b	0.089	69.6
J_T	0.12	0.21
k_A	0.04	0.615
k_{m_α}	0	1.67
m_α^o	0	1.65
δ_A	0.019	0.334
δ_{m_α}	0.024	24.5
δ_{A_β}	0	0.13
k_α^f	0	1098
k_α^b	0	998
k_B	0.144	1.56
k_{m_β}	0.232	985
m_β^o	0	0.243
δ_B	0.0011	0.08
δ_{m_β}	0	2.77
δ_{B_α}	0.031	30.19
k_β^f	4.7	1094
k_β^b	0	14.02
$k_{\beta^R}^f$	0	1003
$k_{\beta^R}^b$	0	959

*Any value in this range will maintain bistability in the model with other parameters held at their baseline values from **Table 2**.*

4.3. Manifolds in Three-Variable Model

We found that our three-variable minimal model retains the bistability and dynamics of the 15-variable model. Therefore, we use the three-dimensional manifolds in this model to better understand the behavior of the full model.

The stable manifold appears to be near-planar in the STAT-APT-SLBO phase space. The UPD value determines the position of the plane, with the manifold shifting from the STAT-SLBO plane to the STAT-APT plane as UPD increases, thus increasing STAT and SLBO production. **Figure 6** shows how the manifold shifts as UPD increases from 0.0133 to 4, values that cover much of the bistable range of UPD.

To visualize the dynamics for each of the three values of UPD in the bistable region used in **Figure 6**, we plot trajectories to demonstrate how the manifold affects the outcomes of different initial conditions. Initial conditions that lie just below the manifold are attracted to the stable manifold near the unstable steady state and then repelled toward the stationary steady state. Initial conditions that lie just above the manifold behave similarly but converge to the motile steady state. **Figure 7** depicts this behavior, with initial conditions that progress to the stationary steady state and the corresponding trajectories in blue and initial conditions that progress to the motile steady state and the corresponding trajectories in red. Three initial conditions and trajectories for each steady state are plotted. The motile steady

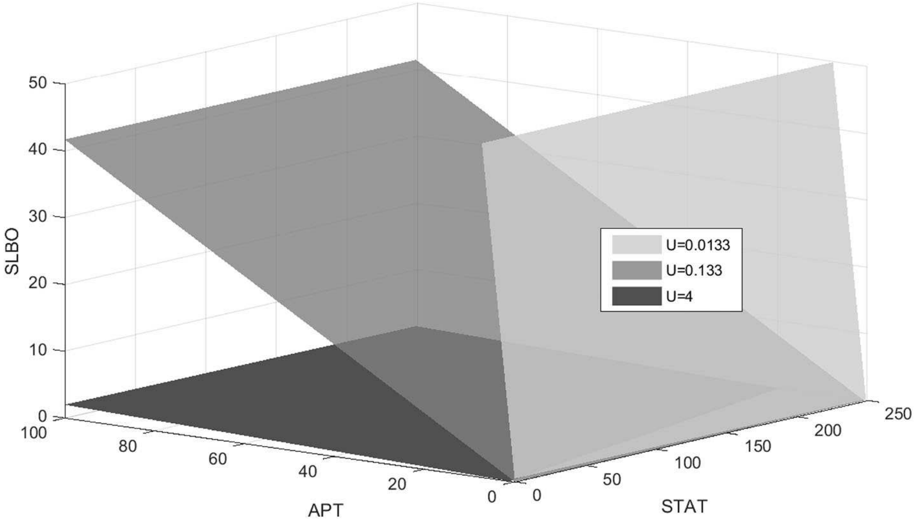


FIGURE 6 | Stable manifold when UPD = 0.0133, UPD = 0.133, and UPD = 4. As UPD increases, the manifold shifts from the STAT-SLBO plane to the STAT-APT plane.

state is in red, the stationary steady state is in blue, and the unstable steady state is in black. The unstable steady state lies on the 2D stable manifold.

When UPD is low, such as in **Figure 7A**, initial conditions in most of the phase space will result in the stationary steady state. When UPD is high, such as in **Figure 7C**, initial conditions in most of the phase space will result in the motile steady state.

Figure 7D combines **Figures 7A–C** to show how the shifting manifold affects the system. As the value of UPD increases and the manifold moves below additional initial conditions, the motile basin of attraction expands and the motile and stationary steady states shift slightly. Movies showing a 3D rotation of each space and time series of each trajectory are available in the **Figures S4–S8**.

4.4. Delay From miRNA

The rate at which APT-induced miRNAs lead to degradation of *stat* mRNA is controlled in the model by the parameter $\delta_{A\sigma}$. **Figure 8A** shows time courses of S in the three-variable model with UPD = 4 for different values of $\delta_{A\sigma}$. We established a typical value of $\delta_{A\sigma} = 0.05$ that allows S to equilibrate around 200 min, which is consistent with the time it takes for border cells to respond to UPD (Starz-Gaiano et al., 2008; Manning et al., 2015). The STAT level when $\delta_{A\sigma} = 0.05$ is well above the threshold needed for the cell to become motile. An increase to $\delta_{A\sigma} = 0.17382$ delays STAT convergence by a significant amount of time, about 600 min. If we increase $\delta_{A\sigma}$ to just 0.18, S never elevates within 1,000 min. Thus, the cell remains stationary. Each time course in **Figure 8** represents a different outcome of the JAK/STAT pathway: GO, where SLBO dominates ($\delta_{A\sigma} = 0.05$); STOP, where APT dominates ($\delta_{A\sigma} = 0.18$); and SLOW, where the transition to motility is delayed ($\delta_{A\sigma} = 0.17382$). The mechanisms of how APT-activated miRNAs affect STAT are still being analyzed, but initial findings support the idea that changes in miRNA activity can cause delays similar to those seen in our model (Yoon et al., 2011; Luo and Sehgal., 2012; Monahan and Starz-Gaiano, 2016; Sun et al., 2015). Delays in STAT activation and failure of activation are possible within a realistic time frame.

For basal levels of $\delta_{A\sigma}$ the system has normal STAT activation. However, as $\delta_{A\sigma}$ is increased the manifold separating the GO and STOP basins of attraction gets close to the initial condition, slowing the rise time to activation. If $\delta_{A\sigma}$ is raised enough the manifold crosses the initial condition and the trajectory is attracted to the STOP fate. **Figure 8B** further illustrates this delay. The 3D plot shows that at $t = 300$ the SLOW cell fate still has very low STAT and SLBO production. At the same time the GO and STOP cell fates have almost converged to their respective steady states. This change in outcome for one initial condition occurs because the increase in $\delta_{A\sigma}$ causes the manifold to shift just above the initial condition. This can be seen in **Figure 9**, where the initial condition lies between the SLOW manifold and the STOP manifold.

4.5. Experimental Tests

We tested our models to confirm if they reproduced certain behaviors identified in various experiments. It has been shown that if STAT activity is blocked by stage 9 of cell migration, the level of APT protein drops by about half (Starz-Gaiano et al., 2008, 2009; Yoon et al., 2011; Monahan and Starz-Gaiano, 2015). It is also known that increasing the initial condition of APT should decrease the level of STAT protein and activity (Starz-Gaiano et al., 2008). These behaviors should be achievable by our models.

The three-variable model (Equations 29–31) reproduces the experimental behavior of APT when STAT is knocked down, as shown in **Figure 10A**. STAT initially converges to a value of about 4. Setting $k_S = 0$ causes STAT to be constantly zero. This causes APT to converge to an equilibrium roughly half

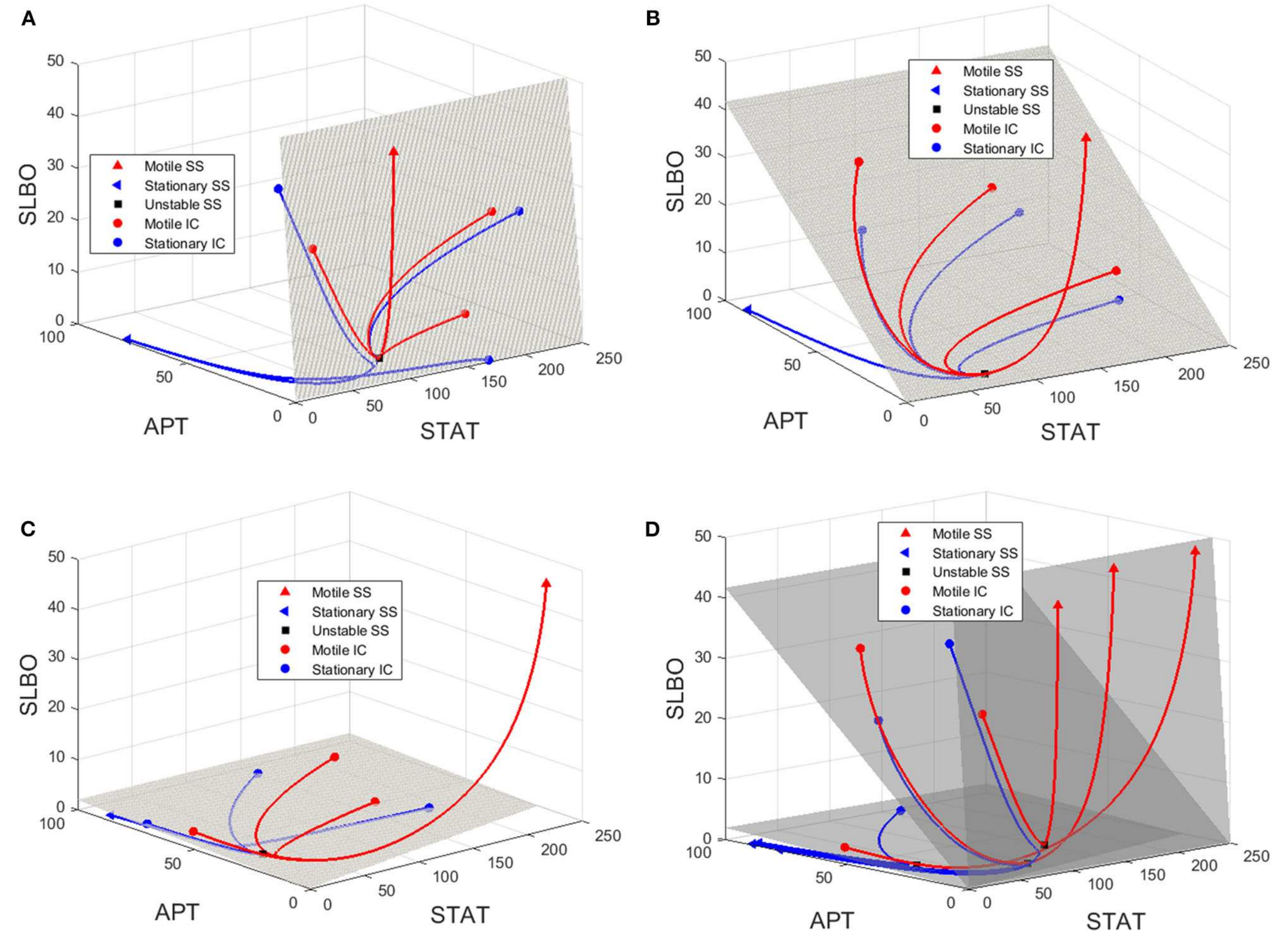


FIGURE 7 | Manifolds and trajectories for three values of UPD. Stationary initial conditions and trajectories are in blue and motile initial conditions and trajectories are in red. The motile steady state is in red, the stationary steady state is in blue, and the unstable steady state is in black. **(A)** Manifold and trajectories for UPD = 0.0133. **(B)** Manifold and trajectories for UPD = 0.133. **(C)** Manifold and trajectories for UPD = 4. Movies showing a 3D rotation of each space and time series of each trajectory are available in the **Figures S4**-**S8**. **(D)** A-C combined to show how the motile basin of attraction expands as UPD increases and the manifold shifts.

that of the stationary steady state. APT acts as an inhibitor to STAT activity, so for higher APT initial conditions we should see lower levels of STAT. This also occurs in the three-variable model (**Figure 10B**).

4.6. Comparison to Previous Models

Previous mathematical models of the JAK/STAT pathway in *Drosophila* have focused on matching the qualitative behavior of the system. We compare our mechanistic model to the models developed in Starz-Gaiano et al. (2008) and Yoon et al. (2011) in **Table 4**. We rewrite our equations and use composite parameters to allow for easier comparison. The names of variables for STAT and SLBO differ across the models.

Our model and the earlier models share the same basic structure. The equations for STAT, APT, and SLBO all have a production term and a degradation term. The production terms include the cross-reactions of the proteins. However, our model describes these interactions in more detail, accounting for more of the known molecular interactions and biochemical production/degradation rates. For example, our STAT equation includes the feedback inhibition from APT. Our model also includes STAT independent production rates for both APT and SLBO, while the earlier models only include this for APT. By including the details of these molecular interactions, we gain the ability to mathematically examine the bistability of the JAK/STAT pathway.

Dynamic UPD is a feature of earlier models. Earlier models also include that levels of active STAT are positively regulated by production of UPD. Our model can be combined with the UPD dynamics developed in Manning et al. (2015). They used a partial differential equation to model the change in UPD concentration over time in three-dimensional extracellular space (Manning et al., 2015). This more detailed version of UPD dynamics includes the spatial element of how UPD diffuses from the polar cells into the border cells. It also helps to explain how border cells farther from the polar cells, which receive less UPD, can be delayed in becoming motile. Additionally, modeling

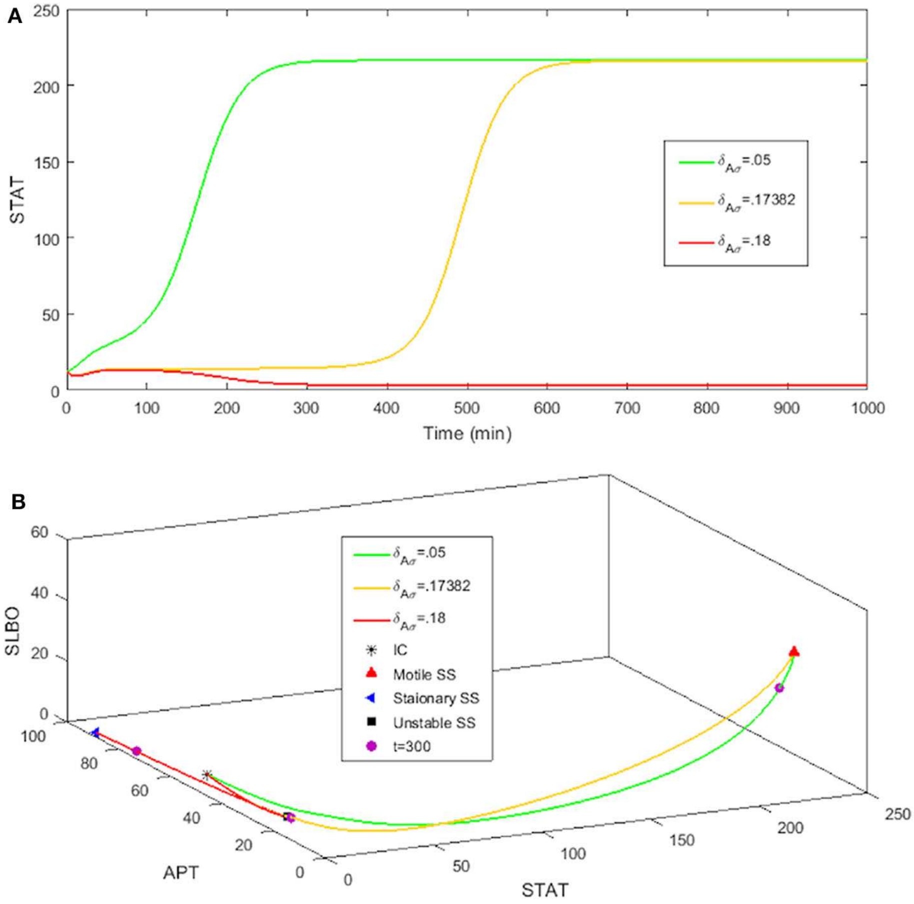


FIGURE 8 | Three levels of miRNA controlled by $\delta_{A\sigma}$: GO $\delta_{A\sigma} = 0.05$, STOP $\delta_{A\sigma} = 0.18$, SLOW $\delta_{A\sigma} = 0.17382$ with initial conditions STAT = 12, APT = 56, SLBO = 1.5, and UPD = 4. **(A)** Time courses of STAT. As $\delta_{A\sigma}$ increases STAT activation is delayed. **(B)** 3D plot of the three cell fates in **(A)**. The motile steady state is in red, the stationary steady state is in blue, and the unstable steady state is in black. $t = 300$ on each trajectory is marked in purple.

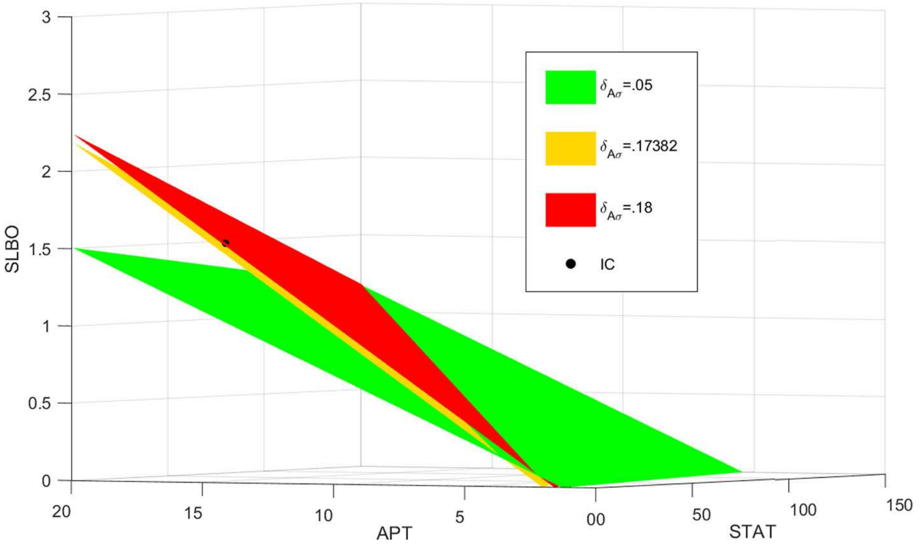


FIGURE 9 | Manifolds for $\delta_{A\sigma} = 0.05$, $\delta_{A\sigma} = 0.18$, $\delta_{A\sigma} = 0.17382$. As $\delta_{A\sigma}$ increases, the manifold shifts above the initial condition in black.

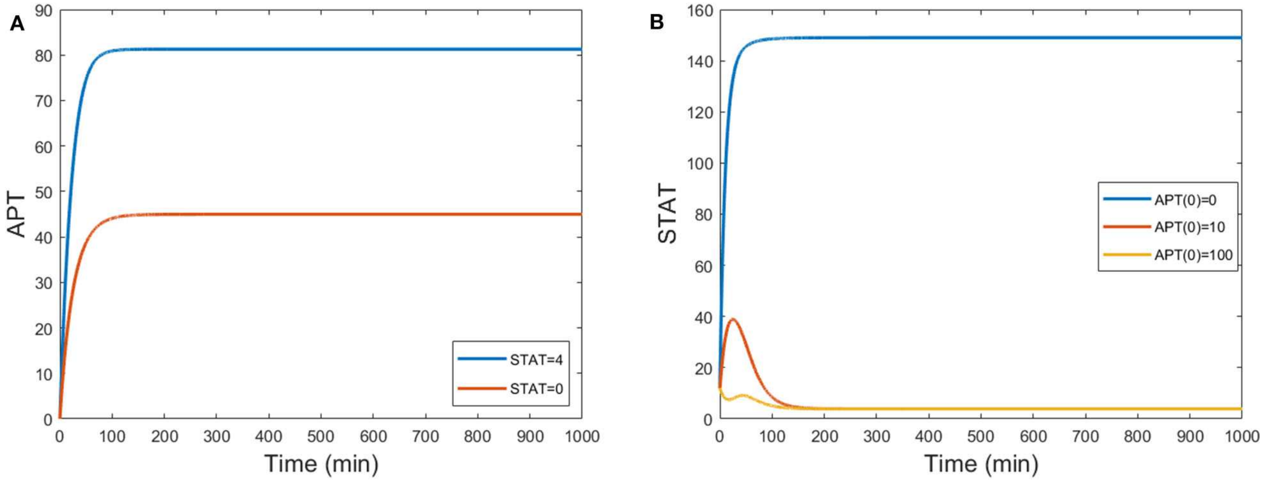


FIGURE 10 | (A) Time courses of APT show levels decrease by half when STAT is knocked down from a steady state value of around 4 to one of 0 by setting $k_S = 0$. **(B)** Time courses of STAT for different initial conditions of APT. When UPD = 0.1, STAT steady state levels decrease for high initial conditions of APT. S(0) = 12 and B(0) = 1.5.

TABLE 4 | Comparison of models.

Variable	Starz-Gaiano et al. (2008)	Yoon et al. (2011)	This paper
JAK/STAT	$\frac{dJ}{dt} = \frac{0.04U}{1+10P^2+0.1J^2+\frac{33A^2}{1+15S}} - 0.01J$	$\frac{dJ}{dt} = \frac{p_jU}{1+q_jP^2+\frac{a_jR}{1+k_jS^2}} - r_jJ$	$\frac{dS}{dt} = \frac{k_S k_{m_\sigma} \frac{S^2}{S^2+k_m^2} \frac{U}{U+K_{UJ}} + m_\sigma^o}{(\delta_1+A)\left[\frac{K_\sigma k_{S_2^*}}{k_{c_1} J_T} + \frac{S^2}{S^2+k_m^2} \frac{U}{U+K_{UJ}}\right]} - \delta_S S$
UPD	$\frac{dU}{dt} = 0.01P - 0.006U + 0.1\frac{d^2U}{dx^2}$	$\frac{dU}{dt} = p_uP - r_uU + d_u\frac{d^2U}{dx^2}$	U is a parameter.
APT	$\frac{dA}{dt} = \frac{0.0001J}{1+4A^2+S} - 0.0001A + 1\text{e-}5$	$\frac{dA}{dt} = \frac{p_aJ}{1+q_aS} - r_aA + b_a$	$\frac{dA}{dt} = \frac{\hat{k_1}}{(\delta_2+B^2)\frac{S^2}{S^2+k_m^2}\frac{U}{U+K_{UJ}}} - \delta_A A + \frac{K_1}{\delta_2+B^2}$
SLBO	$\frac{dS}{dt} = \frac{0.002J^2}{1+0.5A^2} - 0.0005S$	$\frac{dS}{dt} = \frac{p_sJ^2}{1+q_sA} - r_sS$	$\frac{dB}{dt} = \frac{\hat{k_2}\hat{k_3}\frac{S^2}{S^2+k_m^2}\frac{U}{U+K_{UJ}}}{(\delta_3+A)(1+\hat{k_3}\frac{S^2}{S^2+k_m^2}\frac{U}{U+K_{UJ}}+\frac{A}{K_{\beta R}})} - \delta_B B + \frac{K_2}{\delta_3+A}$
miRNA279		$\frac{dR}{dt} = \frac{p_rA}{1+q_rK^2} - r_rR$	
Ken		$\frac{dK}{dt} = p_kJ - r_kK$	

Composite parameters: $K_{UJ} = \frac{k_{UJ}^b}{k_{UJ}^f}, \delta_1 = \frac{\delta_{m_\sigma}}{\delta_{A\sigma}}$,
$\hat{k_1} = \frac{k_A k_{m_\alpha} K_\alpha k_{S_2^*}}{\delta_{B\alpha} k_{c_1} J_T}, K_1 = \frac{k_A m_\alpha^o}{\delta_{B\alpha}}, \delta_2 = \frac{\delta_{m_\alpha}}{\delta_{B\alpha}}, \hat{k_2} = \frac{k_B k_{m_\beta}}{\delta_{A\beta}}, \hat{k_3} = \frac{k_{c_1} J_T}{k_{S_2^*} K_\beta}, K_2 = \frac{k_B m_\beta^o}{\delta_{A\beta}}, \delta_3 = \frac{\delta_{m_\beta}}{\delta_{A\beta}}$.

in Peercy and Starz-Gaiano (2020) gives details of how the geometry of the egg chamber influences cell fate and migration dynamics. They also examine models of velocity and migratory cohort size as well as how clusters of border cells function to guide directional movement. Together our mechanistic model, the spatial dynamics of UPD from Manning et al. (2015), and models of collective migration from Peercy and Starz-Gaiano (2020) help us gain a better understanding of the threshold for motility in border cells.

5. DISCUSSION

We reduced the mechanistic Ge and Stonko model in stages to arrive at our minimal three-variable model. This final model allows us to better understand bistability in the JAK/STAT pathway. Our minimal model retains the parameters from the more complex model, allowing easier analysis but retaining critical properties. We discovered that non-linearity in SLBO repression of *apt* mRNA translation (Equation 9), and to a lesser extent APT repression of the *slbo* gene is required for bistability. The model displays the bistability of the stationary and motile cell states expected from experimental data for medium saturation of STAT activity. It has been shown that modulating UPD expression affects the number of migratory cells (Manning et al., 2015; Van De Bor et al., 2011; Silver and Montell, 2001; Xi et al., 2003; Grammont and Irvine, 2002). If UPD is low there are no migratory cells and if UPD is high there are many. This can induce normally stationary cells far away from the anterior to become motile (Manning et al., 2015).

We established parameters that showed bistability was obtainable under realistic conditions. Every parameter in the 15-variable model was found to have a wide range of values that guaranteed bistability. The robustness with respect to parameter values also suggests the biophysical utility of the model. Cell migration is an essential biological process, so the JAK/STAT pathway necessitates robustness under a range of parameter values, which the model supports. It makes sense that the JAK/STAT pathway would be able to operate successfully under a range of parameter values.

Ge and Stonko assumed that a key aspect necessary for bistability in APT and SLBO is cooperativity in SLBO repressing *apt* mRNA. In our dynamic STAT model, we found that the feedback inhibition of APT on STAT is necessary for bistability in STAT activity coming from the bistability between APT and SLBO. We also found that the initial amount of APT present in a cell is a major factor in whether or not the cell will become motile through affecting the level of STAT activity. Our minimal reduced model also reproduces behavior seen in experiments, such as declines in APT expression when STAT is knocked down and how initial levels of APT affect final levels of STAT (Starz-Gaiano et al., 2008).

We also showed that delays in STAT activation and failure of activation are possible within a realistic time frame. By controlling the degree of feedback inhibition of APT on STAT we can induce a delay in the transition to cell motility or cause the cell to remain stationary. This result is due to the proximity of the UPD level to a limit point bifurcation.

Assumptions made in the development of this model lead to some limitations. Specific parameter values for the exact reactions occurring in the egg chamber are not known, so we performed sensitivity analysis to show the robustness in parameter values. The STAT dynamics were reduced under a number of assumptions. One was the decision to ignore the c_2 equation, which allows APT to act as a buffer on STAT. Since the action of APT on STAT is known to function by preventing activation through miRNAs (Yoon et al., 2011) as well as limiting active protein present (Monahan and Starz-Gaiano, 2015, 2016), this assumption might oversimplify the larger system. Additionally, the quasi-steady state assumptions made in reducing STAT dynamics may also oversimplify the model.

Further study into the various methods by which APT inhibits STAT would enable us to improve how the model captures this interaction. Experiments that change the level of APT or STAT may give more information about how the delay in specification affects the migration of the cell cluster. Experimenting with *apt* mutants could tell us more about how the bistability we show in cell specification ultimately affects cluster migration and the further development of the egg chamber. When STAT is abnormally high more intermediate border cells trail behind the polar cells (Silver and Montell, 2001; Silver et al., 2005; Monahan and Starz-Gaiano, 2015; Starz-Gaiano et al., 2008). Socs36E limits STAT signaling and higher Socs36E reduces motile cell number. Both of these cases may suggest how a delay in cell specification affects cell migration.

It may also be important to consider how APT and SLBO may interact through miRNAs. Currently there is little data beyond the existence of these miRNAs (Ge et al., 2012). miRNA interactions directly affect bistability in the model, so greater detail could improve the model. Experiments that control the level of UPD secreted could identify how quickly the signal enters border cells and the levels of UPD which correspond with the activated STAT threshold to induce motility. To further gauge the STAT levels it would also be valuable to know the level of APT present in border cells through activation by EYA, prior to STAT activation and upregulation by STAT.

The JAK/STAT signaling pathway is known to be well-conserved. Specifically, it seems to be comparable in *Drosophila* and in humans (Trivedi and Starz-Gaiano, 2018; Arbouzova and Zeidler, 2006; Amoyel and Bach, 2012; Amoyel et al., 2014). Thus, as the model we have developed helps to explain cell motility in *Drosophila*, it may prove useful to our understanding of the process in humans. The "decision" for tumor cells to become motile is often a turning point for cancer progression. Additionally, STAT signaling is also well-known for controlling stem cell division decisions. Thus, the decisiveness of STAT-based signaling may have a variety of roles in different cell types, but the model and its bistability shown in this paper may help to explain how this STAT signaling operates in different situations.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

BP and MS-G contributed conception and methodology, and reviewed and edited the manuscript. AB was the main writer the manuscript, and developed the reduced model, performed analysis, and ran simulations. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported in part by National Science Foundation grant IOS-1656550 to MS-G. AB was funded in part by a summer research grant from the Dean of the College of Natural and Mathematical Sciences at UMBC.

ACKNOWLEDGMENTS

We would like to acknowledge the previous work done within the NSF-funded UBM program at UMBC that forms the basis of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00803/full#supplementary-material

REFERENCES

Aman, A., and Piotrowski, T. (2010). Cell migration during morphogenesis. *Dev. Biol.* 341, 20–33. doi: 10.1016/j.ydbio.2009.11.014

Amoyel, M., Anderson, A. M., and Bach, E. A. (2014). JAK/STAT pathway dysregulation in tumors: a Drosophila perspective. *Semin. Cell Dev. Biol.* 28, 96–103. doi: 10.1016/j.semcdb.2014.03.023

Amoyel, M., and Bach, E. A. (2012). Functions of the Drosophila JAK-STAT pathway: lessons from stem cells. *Jak Stat* 1, 176–183. doi: 10.4161/jkst.21621

Arbouzova, N. I., and Zeidler, M. P. (2006). JAK/STAT signalling in Drosophila: insights into conserved regulatory and cellular functions. *Development* 133, 2605–2616. doi: 10.1242/dev.02411

Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. (2007). The microRNA.org resource: targets and expression. *Nucleic. Acids. Res.* 36, 149–153. doi: 10.1093/nar/gkm995

Cai, D., Chen, S.-C., Prasad, M., He, L., Wang, X., Choesmel-Cadamuro, V., et al. (2014). Mechanical feedback through E-cadherin promotes direction sensing during collective cell migration. *Cell* 157, 1146–159. doi: 10.1016/j.cell.2014.03.045

Chen, Y., Kotian, N., Aranjuez, G., Chen, L., Messer, C. L., Burtscher, A., et al. (2020). Protein phosphatase 1 activity controls a balance between collective and single cell modes of migration. *Elife* 9:e52979. doi: 10.7554/eLife.52979.sa2

Ekas, L. A., Cardozo, T. J., Flaherty, M. S., McMillan, E. A., Gonsalves, F. C., and Bach, E. A. (2010). Characterization of a dominant-active stat that promotes tumorigenesis in Drosophila. *Dev. Biol.* 344, 621–636. doi: 10.1016/j.ydbio.2010.05.497

Ermentrout, B. (2002). *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Friedl, P., and Mayor, R. (2017). Tuning collective cell migration by cell–cell junction regulation. *Cold Spring Harbor Perspect. Biol.* 9:a029199. doi: 10.1101/cshperspect.a029199

Ge, X., Stonko, D., Peercy, B., and Starz-Gaiano, M. (2012). Modelling a cellular response to a gradient. *UMBC Rev.* 13, 92–113.

Ghiglione, C., Devergne, O., Georgenthum, E., Carballès, F., Médioni, C., Cerezo, D., et al. (2002). The Drosophila cytokine receptor Domeless controls border cell migration and epithelial polarization during oogenesis. *Development* 129, 5437–5447. doi: 10.1242/dev.00116

Grammont, M., and Irvine, K. D. (2002). Organizer activity of the polar cells during *Drosophila* oogenesis. *Development* 129, 5131–5140. Available online at: https://dev.biologists.org/content/129/22/5131

Guido, N. J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C. R., et al. (2006). A bottom-up approach to gene regulation. *Nature* 439, 856–860. doi: 10.1038/nature04473

Halford, S. E., and Marko, J. F. (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32, 3040–3052. doi: 10.1093/nar/gkh624

Hargrove, J. L., Hulsey, M. G., and Beale, E. G. (1991). The kinetics of mammalian gene expression. *Bioessays* 13, 667–674. doi: 10.1002/bies.950131209

Harris, R. E., Pargett, M., Sutcliffe, C., Umulis, D., and Ashe, H. L. (2011). Brat promotes stem cell differentiation via control of a bistable switch that restricts BMP signaling. *Dev. Cell* 20, 72–83. doi: 10.1016/j.devcel.2010.11.019

Hilton, D., and Nicola, N. (1992). Kinetic analyses of the binding of leukemia inhibitory factor to receptor on cells and membranes and in detergent solution. *J. Biol. Chem.* 267, 10238–10247.

Karsten, P., Plischke, I., Perrimon, N., and Zeidler, M. P. (2006). Mutational analysis reveals separable DNA binding and trans-activation of Drosophila STAT92E. *Cell. Signal.* 18, 819–829. doi: 10.1016/j.cellsig.2005.07.006

Leonard, C. E., and Taneyhill, L. A. (2020). The road best traveled: neural crest migration upon the extracellular matrix. *Semin. Cell Dev. Biol.* 100, 177–185. doi: 10.1016/j.semcdb.2019.10.013

Lewin, B. (2004). *Genes VIII*. Upper Saddle River, NJ: Pearson Prentice Hall.

Lui, P.-Y., Jin, D.-Y., and Stevenson, N. J. (2015). MicroRNA: master controllers of intracellular signaling pathways. *Cell. Mol. Life Sci.* 72, 3531–3542. doi: 10.1007/s00018-015-1940-0

Luo, W., and Sehgal., A. (2012). Regulation of circadian behavioral output via a MicroRNA-JAK/STAT circuit. *Cell* 148, 765–779. doi: 10.1016/j.cell.2011.12.024

Manning, L. A., Weideman, A. M., Peercy, B. E., and Starz-Gaiano, M. (2015). Tissue landscape alters adjacent cell fates during Drosophila egg development. *Nat. Commun.* 6:7356. doi: 10.1038/ncomms8356

McGregor, J. R., Xi, R., and Harrison, D. A. (2002). JAK signaling is somatically required for follicle cell differentiation in Drosophila. *Development* 129, 705–717. Available online at: https://dev.biologists.org/content/129/3/705

Monahan, A., and Starz-Gaiano, M. (2016). Apontic regulates somatic stem cell numbers in Drosophila testes. *BMC Dev. Biol.* 16:5. doi: 10.1186/s12861-016-0103-3

Monahan, A. J., and Starz-Gaiano, M. (2015). Socs36E limits STAT signaling via Cullin2 and a SOCS-box independent mechanism in the Drosophila egg chamber. *Mech. Dev.* 138(Pt 3), 313–327. doi: 10.1016/j.mod.2015.08.003

Montell, D. J., Rorth, P., and Spradling, A. C. (1992). Slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes Drosophila C/EBP. *Cell* 71, 51–62. doi: 10.1016/0092-8674(92)90265-E

Montell, D. J., Yoon, W. H., and Starz-Gaiano, M. (2012). Group choreography: mechanisms orchestrating the collective movement of border cells. *Nat. Rev. Mol. Cell Biol.* 13, 631–645. doi: 10.1038/nrm3433

Nicholson, S. E., and Nicola, N. A. (eds.). (2013). *JAK-STAT Signalling: Methods and Protocols*, Volume 967 of *Methods in Molecular Biology*. New York, NY: Humana Press.

Olson, H. M., and Nechiporuk, A. V. (2018). Using zebrafish to study collective cell migration in development and disease. *Front. Cell Dev. Biol.* 6:83. doi: 10.3389/fcell.2018.00083

Parsaeian, A., de la Cruz, M. O., and Marko, J. F. (2013). Binding-rebinding dynamics of proteins interacting non-specifically with a long DNA molecule. *Phys. Rev. E Stat Nonlinear Soft Matter Phys.* 4:040703. doi: 10.1103/PhysRevE.88.040703

Peercy, B. E., and Starz-Gaiano, M. (2020). Clustered cell migration: modeling the model system of Drosophila border cells. *Semin. Cell Dev. Biol.* 100, 167–176. doi: 10.1016/j.semcdb.2019.11.010

Rorth, P. (1994). Specification of C/EBP function during Drosophila development by the bZIP basic region. *Science* 266, 1878–1881. doi: 10.1126/science.7997882

Rorth, P., Szabo, K., and Texido, G. (2000). The level of C/EBP protein is critical for cell migration during Drosophila oogenesis and is tightly controlled by regulated degradation. *Mol. Cell* 6, 23–30. doi: 10.1016/S1097-2765(05)00008-0

Saadin, A., and Starz-Gaiano, M. (2016). Circuitous genetic regulation governs a straightforward cell migration. *Trends Genet.* 32, 660–673. doi: 10.1016/j.tig.2016.08.001

Silver, D. L., Geisbrecht, E. R., and Montell, D. J. (2005). Requirement for JAK/STAT signaling throughout border cell migration in Drosophila. *Development* 132, 3483–3492. doi: 10.1242/dev.01910

Silver, D. L., and Montell, D. J. (2001). Paracrine signaling through the JAK/STAT pathway activates invasive behavior of ovarian epithelial cells in Drosophila. *Cell* 107, 831–841. doi: 10.1016/S0092-8674(01)00607-9

Starz-Gaiano, M., Melani, M., Meinhardt, H., and Montell, D. (2009). Interpretation of the UPD/JAK/STAT morphogen gradient in Drosophila follicle cells. *Cell Cycle* 8, 2917–2925. doi: 10.4161/cc.8.18.9547

Starz-Gaiano, M., Melani, M., Wang, X., Meinhardt, H., and Montell, D. J. (2008). Feedback inhibition of JAK/STAT signaling by apontic is required to limit an invasive cell population. *Dev. Cell* 14, 726–738. doi: 10.1016/j.devcel.2008.03.005

Stuelten, C. H., Parent, C. A., and Montell, D. J. (2018). Cell motility in cancer invasion and metastasis: insights from simple model organisms. *Nat. Rev. Cancer* 18, 296–312. doi: 10.1038/nrc.2018.15

Sun, K., Jee, D., de Navas, L. F., Duan, H., and Lai, E. C. (2015). Multiple *in vivo* biological processes are mediated by functionally redundant activities of Drosophila mir-279 and mir-996. *PLoS Genet.* 11:e1005245. doi: 10.1371/journal.pgen.1005245

Thurmond, J., Goodman, J., Strelets, V., Attrill, H., Gramates, L., Marygold, S., et al. (2019). Flybase 2.0: the next generation. *Nucleic Acids Res.* 47, D759–D765. doi: 10.1093/nar/gky1003

Trivedi, S., and Starz-Gaiano, M. (2018). Drosophila Jak/STAT signaling: regulation and relevance in human cancer and metastasis. *Int. J. Mol. Sci.* 19:4056. doi: 10.3390/ijms19124056

Van De Bor, V., Zimniak, G., Cerezo, D., Schaub, S., and Noselli, S. (2011). Asymmetric localisation of cytokine mRNA is essential for JAK/STAT activation during cell invasiveness. *Development* 138, 1383–1393. doi: 10.1242/dev.056184

Ward, L., Howlett, G., and Hammacher, A. (1995). Use of a biosensor with surface plasmon resonance detection for the determination of binding constants: measurement of interleukin-6 binding to the soluble interleukin-6 receptor. *Biochemistry* 34, 2901–2907. doi: 10.1021/bi00009a021

Wright, V. M., Vogt, K. L., Smythe, E., and Zeidler, M. P. (2011). Differential activities of the Drosophila JAK/STAT pathway ligands Upd, Upd2 and Upd3. *Cell. Signal.* 23, 920–927. doi: 10.1016/j.cellsig.2011.01.020

Xi, R., McGregor, J. R., and Harrison, D. A. (2003). A gradient of JAK pathway activity patterns the anterior-posterior axis of the follicular epithelium. *Dev. Cell* 4, 167–177. doi: 10.1016/S1534-5807(02)00412-4

Yang, E., Henriksen, M. A., Schaefer, O., Zakharova, N., and Darnell, J. E. (2002). Dissociation time from DNA determines transcriptional function in a STAT1 linker mutant. *J. Biol. Chem.* 277, 13455–13462. doi: 10.1074/jbc.M112038200

Yoon, W. H., Meinhardt, H., and Montell, D. J. (2011). miRNA-mediated feedback inhibition of JAK/STAT morphogen signalling establishes a cell fate threshold. *Nat. Cell Biol.* 13, 1062–1069. doi: 10.1038/ncb2316

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 31 July 2020
doi: 10.3389/fphys.2020.00848

Awakening the HSC: Dynamic Modeling of HSC Maintenance Unravels Regulation of the TP53 Pathway and Quiescence

Nensi Ikonomi†, Silke D. Kühlwein†, Julian D. Schwab† and Hans A. Kestler*

Institute of Medical Systems Biology, Ulm University, Ulm, Germany

Hematopoietic stem cells (HSCs) provide all types of blood cells during the entire life of the organism. HSCs are mainly quiescent and can eventually enter the cell cycle to differentiate. HSCs are maintained and tightly regulated in a particular environment. The stem cell niche regulates dormancy and awakening. Deregulations of this interplay can lead to hematopoietic failure and diseases. In this paper, we present a Boolean network model that recapitulates HSC regulation in virtue of external signals coming from the niche. This Boolean network integrates and summarizes the current knowledge of HSC regulation and is based on extensive literature research. Furthermore, dynamic simulations suggest a novel systemic regulation of TP53 in homeostasis. Thereby, our model indicates that TP53 activity is balanced depending on external stimulations, engaging a regulatory mechanism involving ROS regulators and RAS activated transcription factors. Finally, we investigated different mouse models and compared them to *in silico* knockout simulations. Here, the model could recapitulate *in vivo* observed behaviors and thus sustains our results.

Keywords: hematopoietic stem cell, Boolean network, dynamic modeling, TP53, stem cell awakening, maintenance of quiescence, niche interactions, homeostasis

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
Nathan Weinstein,
Universidad Nacional Autónoma de México, Mexico
Alfredo Rodríguez,
National Institute of Pediatrics, Mexico

****Correspondence:***
Hans A. Kestler
hans.kestler@uni-ulm.de

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *30 April 2020*
Accepted: *24 June 2020*
Published: *31 July 2020*

Citation:
Ikonomi N, Kühlwein SD, Schwab JD and Kestler HA (2020) Awakening the HSC: Dynamic Modeling of HSC Maintenance Unravels Regulation of the TP53 Pathway and Quiescence. Front. Physiol. 11:848. doi: 10.3389/fphys.2020.00848

1. INTRODUCTION

The hematopoietic system is composed of heterogenic populations. These populations comprise highly specialized cells with unique and peculiar functions. All these different blood cells, approximately $4-5 \cdot 10^{11}$ in total, are produced hierarchically from a population of cells called hematopoietic stem cells (HSCs) (Kaushansky, 2006; Jagannathan-Bogdan and Zon, 2013). Nevertheless, recent results coming from single-cell RNA sequencing suggest that hematopoiesis is not necessarily hierarchical, but can also be radial (Macaulay et al., 2016; Athanasiadis et al., 2017; Laurenti and Göttgens, 2018; Zhang et al., 2018; Yokota, 2019).

HSCs are a particular population of cells characterized by the ability to self-renew and to generate differentiated blood cells (Pietras et al., 2011; Sun et al., 2014; Busch et al., 2015; Sawai et al., 2016). Under steady-state conditions, HSCs are maintained as a quiescent population of cells that can be activated in the presence of external differentiation stimuli (Fleming et al., 1993; Bradford et al., 1997; Cheshier et al., 1999; Passegué et al., 2005). This dormant state is believed to contribute to the long-term preservation of HSCs, mostly by minimizing replication and metabolic activity (Wilson et al., 2009; Warr et al., 2011; Bakker and Passegué, 2013). Fine-tuning of dormancy and the awakening mechanism of HSCs is of crucial relevance for correct hematopoiesis. In fact, not responsive dormant HSCs would lead to hematopoietic failure due to a lack of differentiated blood cells (Wilson et al., 2009). On the other hand, highly active HSCs would get to exhaustion of the population and lack of long-term maintenance of the hematopoietic system (Wilson et al., 2009).

Leading intrinsic players regulating HSC quiescence are polycomb ring finger protein (BMI1), TP53, reactive oxygen species (ROS) regulators and protein kinase B (AKT) signaling (Abbas et al., 2011; Pietras et al., 2011; Suda et al., 2011; Warr et al., 2011). Besides intrinsic factors, HSCs are regulated by the niche. The niche is defined as the specialized bone marrow environment where the HSCs reside (Schofield, 1978). Classically, two types of niches have been individuated for HSCs: the endosteal and the vascular niches. The endosteal niche is responsible for the maintenance of quiescence and is composed mainly of osteoblastic cells. In this context, hypoxic condition and signals promoting quiescence such as transforming growth factor beta (TGF-β) set HSCs in a dormant state (Zhang et al., 2003; Kopp et al., 2005; Sugiyama et al., 2006; Geiger et al., 2007; Jones and Wagers, 2008; Wagner et al., 2008; Guerrouahen et al., 2011). On the other hand, the vascular niche is characterized by endothelial and stromal cells and creates a milieu that supports proliferation, differentiation and trans-endothelial migration of HSCs (Zhang et al., 2003; Kopp et al., 2005; Sugiyama et al., 2006; Geiger et al., 2007; Jones and Wagers, 2008; Wagner et al., 2008; Guerrouahen et al., 2011). Here, a higher oxygen concentration and proliferating signals such as stem cell factor (SCF), fibroblast growth factor (FGF), and other growth factors are mostly present. Overall, these signals stimulate HSC activation and cycling (Coskun and Hirschi, 2010; Lilly et al., 2011). Moreover, HSCs and progenitors can be defined by surface markers as LSK (Lin^- Sca^+ cKit^+) population (Spangrude et al., 1988). Additionally, HSCs can be further divided into long-term HSCs (LT-HSC) and short-term HSCs (ST-HSC) depending again on marker selection. In this context, LT-HSCs are defined as Lin^- Sca^+ $\text{cKit}^+\text{CD150}^+\text{CD48}^+$ (Adolfsson et al., 2001; Kiel et al., 2005; Wilson and Trumpp, 2006; Challen et al., 2009; Foudi et al., 2009; Wilson et al., 2009). Phenotypically these two types of HSCs are distinguished based on their quiescence, long-term self-renewal ability, and engraftment. Hence, even the dormant state itself is regulated distinguishing between highly dormant and more active HSCs (Adolfsson et al., 2001; Kiel et al., 2005; Wilson and Trumpp, 2006; Challen et al., 2009; Foudi et al., 2009; Wilson et al., 2009).

Different models to unravel HSC fate commitment are available (Glauche et al., 2011; Krumsiek et al., 2011; Hamey et al., 2017; Olariu and Peterson, 2019). However, as it was underlined, HSC function does not only reside in its ability to differentiate. Unraveling mechanistic strategies by which HSCs are kept in quiescence and activeness is of crucial interest. Accordingly, the deregulation of HSC function is linked to cancerous conditions and aging (Warr et al., 2011; Bakker and Passegué, 2013). Even if a lot of effort has been implied in understanding the maintenance of HSCs, still many regulatory mechanisms are unknown. Crucial regulators of quiescence such as TP53, BMI1, ataxia-telangiectasia mutated (ATM) have been individuated. However, it is still not clear how they are differentially regulated in dormant (LT-HSC) and more active HSC (ST-HSC) (Warr et al., 2011; Bakker and Passegué, 2013). The regulation of these factors is often believed to be a result of crosstalk among different intrinsic and extrinsic pathways involved in HSC maintenance (Warr et al., 2011; Bakker and Passegué, 2013). Of interest in this context is the regulation of TP53.

TP53 is a ubiquitously expressed protein described to be activated in the context of DNA damage response and regulates cell survival (Toledo and Wahl, 2006; Vousden and Lane, 2007; Meek, 2009). In normal conditions, instead, TP53 is down-regulated by mouse double minute 2 homolog (MDM2) that causes its degradation (Jones et al., 1995; Montes de Oca Luna et al., 1995; Ringshausen et al., 2006). MDM2 expression is further sustained by TP53, causing negative feedback on its activity (Jones et al., 1995; Montes de Oca Luna et al., 1995; Ringshausen et al., 2006). Nevertheless, high levels of TP53 have been detected in HSCs. Its expression is linked to the maintenance of quiescence by induction of growth factor independent one transcriptional repressor (GFI1) (Forsberg et al., 2005; Liu Y. et al., 2009). How both intrinsic and extrinsic factors regulate TP53 in steady-state conditions is still an open question and a central point in further understanding of HSC maintenance (van Os et al., 2009; Pant et al., 2012).

To get new mechanistic insights about HSC maintenance and, in particular, on TP53 regulation, we constructed a Boolean network model describing quiescence and activation of HSCs.

Among dynamic modeling approaches, Boolean network models are used to describe the qualitative dynamic behavior of gene and protein regulatory networks (Kauffman, 1969; Hopfensitz et al., 2013). Boolean network models rely on the assumption of a switch-like behavior for each node (corresponding to a gene or a protein) that can be considered either active (1) or inactive (0) (Kauffman, 1969; Hopfensitz et al., 2013). Moreover, interactions are qualitatively described and summarized in Boolean functions, which do not require frequently unavailable kinetic parameters to be constructed. This peculiarity makes Boolean networks a powerful tool to describe large biological networks (Kauffman, 1969; Hopfensitz et al., 2013). Despite their simplicity, Boolean network models can faithfully represent the behavior of biological networks. In this context, they have been successfully applied to recapitulate different biological conditions such as cancer, aging, and senescence (Herrmann et al., 2012; Dahlhaus et al., 2016; Meyer et al., 2017; Siegle et al., 2018).

In the present work, we performed an extensive literature investigation to construct a Boolean network model recapitulating the regulation of HSC quiescence and cell cycle entry in the dependency on niche stimulation. Our aim is to unravel new mechanistic insights in the maintenance of hematopoietic stem cell quiescence, with particular attention to TP53 regulation. Strikingly, our model recapitulates the different statuses of the HSC, successfully describing dormant LT-HSC, activated ST-HSC, and cycling HSC as well as the progression from one to the other. Moreover, we could suggest a new general mechanism for TP53 regulation in homeostatic conditions. Finally, we tested our regulatory model by assessing its ability to recapitulate different knockout (K.O.) conditions.

Here, the results support our model-based hypothesis of HSCs regulation.

2. MATERIALS AND METHODS

2.1. Boolean Networks

Boolean networks are dynamic mathematical models to describe gene regulatory processes (Kauffman, 1969). These networks are defined as a set of n variables $X = \{x_1, x_2, \ldots, x_n\}, x_i \in \mathbb{B}$ and a corresponding set of transition functions $F = \{f_1, f_2, \ldots., f_n\}, f_i : \mathbb{B}^n \rightarrow \mathbb{B}$. Boolean functions integrate information about regulatory interactions coming from literature statements or other data sources. Here, regulatory dependencies are expressed using logical operators (Kauffman, 1969). Each compound x_i has its own Boolean function f_i which defines whether x_i is present or not. The state of the network at a specific point in time t is defined by a vector $\vec{x}(t) = (x_1(t), \ldots, x_n(t))$ containing the assignments of all compounds x_i within the network at this particular time step. Considering all possible combinations of assignments, this leads to a total number of 2^n possible states. Using synchronous updates, transitions from states at one point in time to their successors, $\vec{x}(t) \mapsto \vec{x}(t+1)$, are computed by updating all compounds at the same time. The transition of each compound is done by applying the corresponding Boolean function $x_i(t+1) = f_i(\vec{x}(t)), f_i : \mathbb{B}^n \rightarrow \mathbb{B}$ (Kauffman, 1969). Due to the deterministic nature of the state space, with synchronous updating, the model eventually enters a recurrent sequence of states called attractor (Kauffman, 1969). Attractors denote the long-term behavior of a system, and in a biological context, they are often related to phenotypes (Kauffman, 1993; Thomas and Kaufman, 2001). All states leading to the same attractor are part of its so-called basin of attraction (Hopfensitz et al., 2013; Schwab et al., 2020).

2.2. Model Setup

For the construction of the Boolean network model describing the maintenance of HSC in their niche, we performed extensive literature research collecting published papers from NCBI and Google Scholar. We collected data in the context of hematopoietic stem cell regulation. Moreover, the influences of niche factors on intrinsic regulators were considered. In the present work, studies from mouse models as well as from human HSCs when available were used. For all components of the model, different levels of regulation have been considered. Detailed descriptions about all interactions and resulting Boolean network functions can be found in **Table 1**. The java framework ViSiBooL was used for modeling (Schwab et al., 2017).

2.3. Model Simulation

The analysis of the dynamic behavior of the HSC model was performed by synchronous updates with the R package BoolNet (Müssel et al., 2010). Simulations were performed by applying the SAT-based attractor search algorithm.

First, analyses of the general network dynamics were performed by exhaustive attractor-search in the entire state space. Resulting attractors were compared to described HSC phenotypes. Here, one attractor was excluded from further analyses due to a complete lack of external stimulation. This does not correspond to a biologically plausible condition. In a further step, progressions from one attractor to the next were studied. We used the attractor pattern of each HSC phenotype as a starting state and altered the external niche stimuli to trigger the cascade. Finally, paths from one attractor to the next were examined. Altogether, we analyzed trajectories from:

- LT-HSC to ST-HSC by activating external cycling stimuli
- ST-HSC to cycling HSC by inactivating external quiescence stimuli
- LT-HSC to cycling HSC by combining activation of external cycling stimuli and inactivation of quiescence stimuli.

We simulated all possible single-compound perturbations (knockout and overexpression, see **Supplementary Section 5**). To validate the model, *in silico* intervention studies were performed and compared to known mouse phenotypes. Knockouts were obtained by fixing the state of desired node(s) constantly to 0. Dynamic analyses of the considered interventions were simulated exhaustively and attractors were analyzed.

All scripts to run the simulations are provided on github (https://github.com/sysbio-bioinf/HSC-boolean-network-model) and reported in the **Supplementary Section 6**.

2.4. Robustness Analysis

We applied stochastic noise to our Boolean network model to evaluate the robustness—and thus the significance—of our model and the applied simulations. First, the ability of the network to compensate noise has been computed based on its attractors and their robustness against noise in the form of random bit flips in their basin of attraction. In our setup, we created a set of 1 million random initial states and computed the attractors for all of these states individually. Next, we created a perturbed copy of each drawn initial state and applied attractor search on all these perturbed states. For perturbation, we used random bit flips (assignments of selected nodes were toggled for 1 to 0 and vice versa). Flipping nodes in the networks' states is a common approach to simulate noise in Boolean networks (see, Qu et al., 2002; Aldana and Cluzel, 2003; Kauffman et al., 2004). To evaluate the robustness of the network, we compared the resulting attractor of each random state and its perturbed copy. Consequently, we could assess if this perturbation causes a switch of the attractors toward different basins. This procedure was exerted with toggling a different amount of random nodes to simulate an increasing amount of noise. We ran the simulation repeatedly, flipping 1, 2, and 3 nodes of each random state. For each number of bit flips, we repeated the simulation three times.

As a second measure for robustness, we analyzed the impact of noise in the model and compared it with randomly generated Boolean networks. Again, random bit flips are applied by adding noise to the network. Then, the corresponding successor state of the original one as well as of the perturbed state is computed. Then, the distance between the two successor states is measured using the normalized Hamming distance. The Hamming distance measures the number of differing bits across two state vectors. It can be formally described as $H(x, y) = \sum_i^n x_i \oplus y_i, \forall x, y \in \mathbb{B}^n$. The normalization is done by dividing the measured Hamming

TABLE 1 | Boolean functions for the HSC model.

Node	Boolean function
External quiescence	External quiescence
External cycling	External cycling
PI3K	RAS
TSC1/2	¬AKT
mTORC1	¬TSC1/2
FOXO3A	External quiescence ∧ ¬AKT
ATM	FOXO3A
Mitochondria	mTORC1
ROS	Mitochondria ∨ ¬ATM ∨ ¬ FOXO3A ∨ ¬BMI1 ∨ ¬ TP53
Autophagy	FOXO3A ∧ ROS ∧ ¬ mTORC1
RAS	External cycling
ETS	RAS ∧ ¬ MEF
MEF	RAS
GSK3β	¬AKT
CTNNB1	¬GSK3β
cMYC	CTNNB1 ∧ ¬GSK3β
BMI1	cMYC ∨ (FOXO3A ∧ ATM)
MDM2	(TP53 ∨ MEF) ∧ ¬CDKN2D ∧ ¬ATM
TP53	¬MDM2
CDKN1C	External quiescence ∨ FOXO3A
CDKN1A	(TP53 ∨ FOXO3A ∨ External quiescence ∨ GFI1) ∧ ¬cMYC
CDKN1B	FOXO3A
GFI1	TP53
RB	¬CCND1 ∧ ¬CCNE1
E2F	¬RB ∧ ¬ GFI1
CCND1	¬CDKN2A ∧ ¬ CDKN1C ∧ cMYC
CCNE1	¬CDKN1C ∧ ((¬CDKN1A ∧ ¬CDKN1B) ∨ CCND1) ∧ E2F
S-phase	E2F ∧ CCNE1
AKT	PI3K
CDKN2A	(ETS ∨ ROS) ∧ ¬BMI1
CDKN2D	(E2F ∨ ROS) ∧ ¬BMI1
Pro-apoptotic proteins	ROS ∧ TP53 ∧ ¬AKT
Anti-apoptotic proteins	(RAS ∨ External quiescence) ∧ ¬GSK3β
CYCS	Pro-apoptotic proteins ∧ ¬Anti-apoptotic proteins
Apoptosis	CYCS ∧ ¬AKT
Senescence	(CDKN2A ∧ ROS) ∨ (TP53 ∧ ROS ∧ CDKN1A)

Compounds are abbreviated according to accepted nomenclature. Regulatory interactions are summarized by logical connectives AND (∧), OR (∨), and NOT (¬). References for each Boolean function are provided in the ***Supplementary Section 1.1****.*

distance by the number of bits in the vectors, $HD(x,y) = \frac{H(x,y)}{n}, \forall x, y \in \mathbb{B}^n$. The distance is an indicator of the ability of the Boolean network to maintain its functionality under noisy conditions. A normalized Hamming distance of zero indicates that the mutation has no effect on evaluated network behavior. The procedure was repeatedly done for 1,000 randomly drawn states. The number of bit flips for the perturbation was set to one. Finally, results were compared to the ones obtained by 1,000 randomly generated Boolean networks. For this computerintensive test, we computed a p-value to evaluate if our null hypothesis "the hamming distance for the constructed network is larger or equal than the distance for the random networks" can be rejected, as follows : $\frac{\sum_i^r (HD_{rnd_i} > HD_c)}{r}$. Here HD_{rnd_i} represents the mean Hamming distance over all states of one random network rnd_i and r is the total number of random networks tested. HD_c represents the mean normalized Hamming distance over all measured states of the constructed network. We considered $p < 0.001$ to be significant. This test procedure is included in the BoolNet package (Müssel et al., 2010). The final aim of this analysis is to test the robustness of the constructed Boolean network model compared to a number of randomly generated ones.

3. RESULTS

3.1. Modeling the HSC Phenotypes

The HSC regulation in the niche is of crucial relevance for the maintenance of the hematopoietic system. To unravel HSC maintenance mechanisms in homeostatic conditions, we constructed a Boolean network model based on extensive literature research (**Table 1**). A version of **Table 1** endowed with references is available in our **Supplementary Section 1.1**.

Here we considered primary intrinsic regulators of HSC quiescence and niche external stimuli involved in their regulation. As intrinsic regulatory factors, we included cell cycle regulators (such as BMI1, MYC protoncogene (cMyc), TP53 together with cyclin-dependent kinases inhibitors (CDKN) and cyclins), proliferative pathways involving phosphoinositide 3-kinase (PI3K) and rat sarcoma (RAS), regulators of ROS, regulators of apoptosis and all potential cross-regulation among them. Furthermore, we also included a node describing autophagy in light of possible enlargement of our work toward aging studies (Ho et al., 2017). Extrinsic stimuli from the niche have been summarized as quiescence signals able to induce cell cycle inhibitors, or as cycling signals activating the RAS/PI3K axis. In total, the network consists of 37 nodes and 78 regulatory interactions (**Figure 1**).

First, we exhaustively simulated our model, obtaining four single state attractors. The network contains two external inputs (external quiescence and external cycling), resulting in four possible combinations. In accordance, each reached attractor shows a different combination of these nodes (**Figure 2**). Given that these input nodes represent niche stimuli, a condition with any stimulation from the environment is considered not realistic. Therefore, we do not consider this condition for further analyses.

Attractor 1 is affected by quiescent external stimulation, while external cycling factors are absent. This attractor shows neither activation of the cell cycle [synthesis phase (S-phase) and cyclins] nor active metabolism or cell growth [mammalian target of rapamycin complex 1 (mTORC1) and ROS]. Instead, many inhibitors of the cell cycle [cyclin-dependent kinase inhibitor 1C (CDKN1C), cyclin-dependent kinase inhibitor 1A (CDKN1A), cyclin dependent kinase inhibitor 1B (CDKN1B), GFI1], as well as ROS and mTORC1 signaling [Forkhead box O3 (FOXO3A), ATM, Tuberous sclerosis protein 1/2 (TSC1/2)] are active.

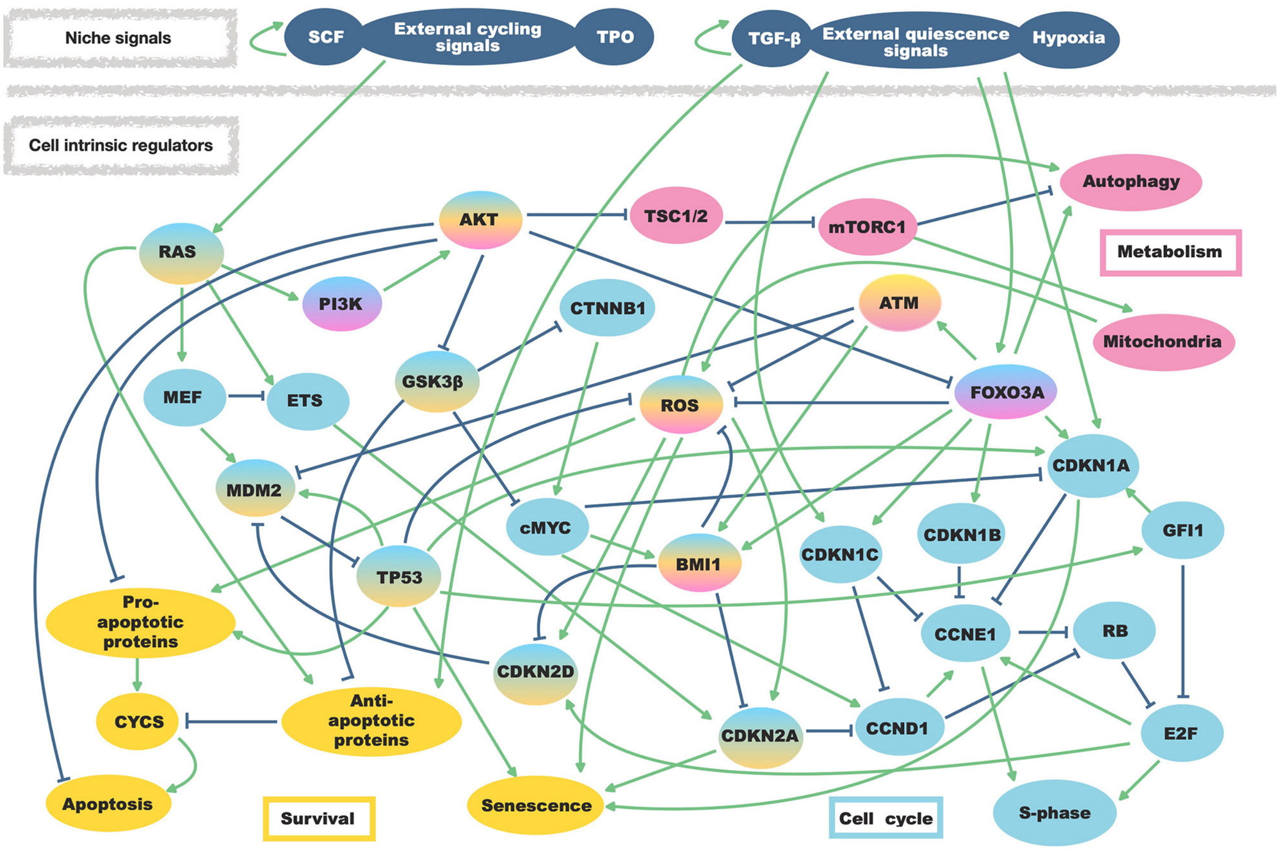
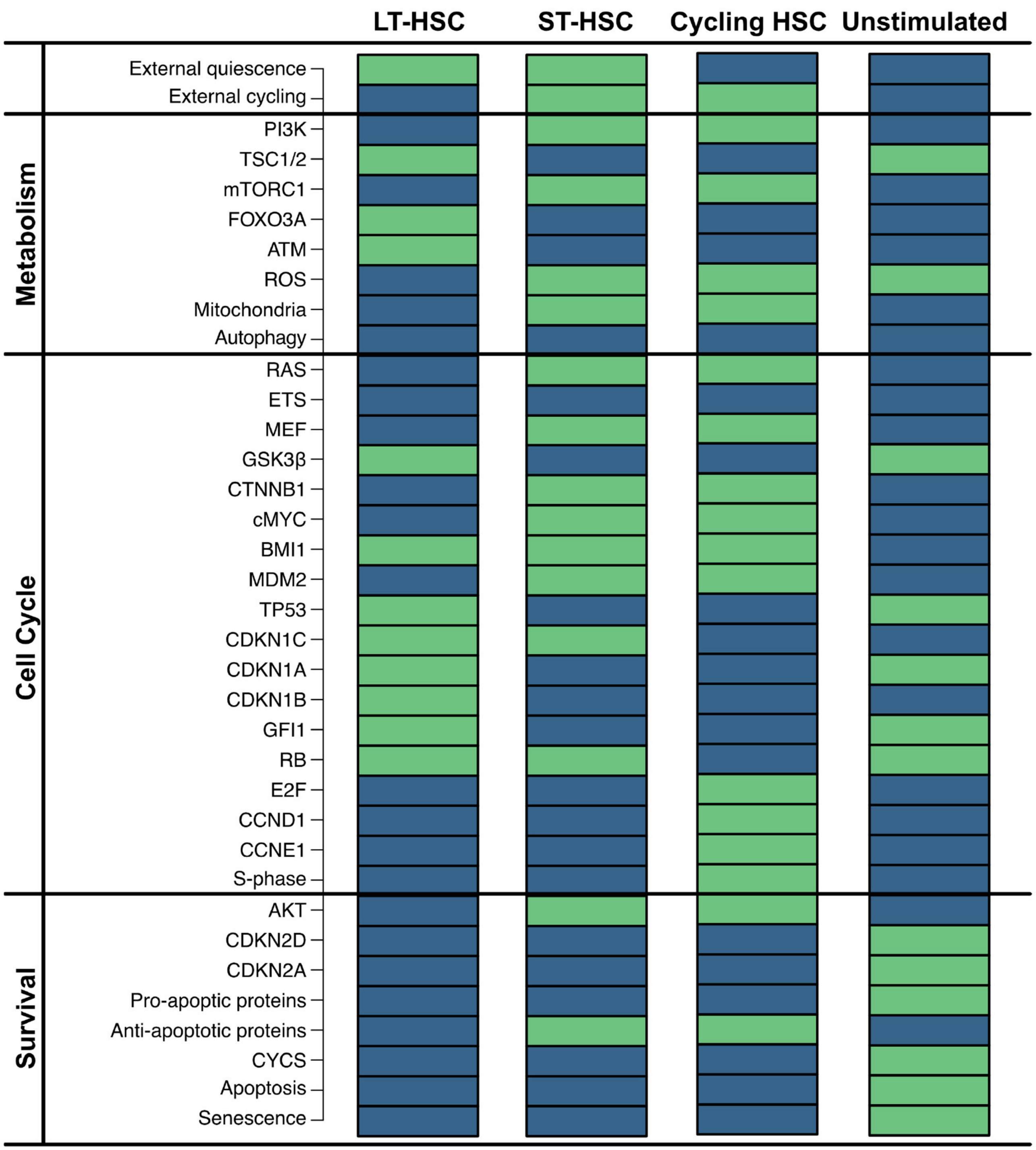


FIGURE 1 | Regulatory interactions of the HSC model. External niche factors as well as regulations within the hematopoietic stem cells are summarized in one model. Regulatory activating interactions are depicted by arrows while bar-headed lines represent inhibiting regulations. Affected mechanisms and involved proteins are separated by colors.

Furthermore, no apoptosis is observable, although TP53 is active. However, pro-apoptotic proteins and ROS that are required for apoptosis induction are inactive. This attractor pattern matches the behavior of a dormant LT-HSC which is considered as a reservoir of the HSCs and typically found in the endosteal region (Kopp et al., 2005; Sugiyama et al., 2006; Geiger et al., 2007; Jones and Wagers, 2008; Wagner et al., 2008; Guerrouahen et al., 2011). The attractor pattern was compared to the expression levels of quiescent HSCs. Here, low ROS (Chen et al., 2008, 2009; Ludin et al., 2014), AKT/mTORC1 (Jang and Sharkis, 2007; Chen et al., 2008, 2009; Ludin et al., 2014; Rodgers et al., 2014; Cabezas-Wallscheid et al., 2017; Baumgartner et al., 2018), RAS (Geest and Coffer, 2009), and cMYC (Wilson et al., 2004; Forsberg et al., 2005; Cabezas-Wallscheid et al., 2017) have also been observed. Additionally, quiescence is described to be maintained by expression of FOXO3A and ATM (Suda et al., 2011; Bakker and Passegué, 2013; Ludin et al., 2014) as well as TP53 activity (Forsberg et al., 2005; Liu Y. et al., 2009; Pant et al., 2012).

Attractor 2 is affected by the activity of both external stimuli. Similarly to attractor 1, the cell cycle (S-phase) is not active based on these external conditions. Hence, this attractor also shows quiescent behavior. However, this time RAS/PI3K signaling as well as mTORC1, ROS, and cMYC are active. Moreover, in attractor 2, many cell cycle inhibitors (CDKN1A, CDKN1B, GFI1) and ROS regulators (FOXO3A and ATM) are inactive. In comparison to attractor 1, apoptosis is still inactive. This might be due to inactive TP53 and the activation of the anti-apoptotic protein. Attractor 2 resembles the phenotype of an ST-HSC. The observed attractor pattern is comparable with the expression levels of ST-HSC (Wilson et al., 2004; Forsberg et al., 2005; Jang and Sharkis, 2007; Chen et al., 2008, 2009; Geest and Coffer, 2009; Liu Y. et al., 2009; Ludin et al., 2014; Rodgers et al., 2014; Cabezas-Wallscheid et al., 2017; Baumgartner et al., 2018). Moreover, in our attractor, the main cell cycle inhibitor CDKN1C is still active. It is also consistent with experimental observations indicating CDKN1C as the crucial regulator of quiescence in HSC (Umemoto et al., 2005; Matsumoto et al., 2011; Tesio and Trumpp, 2011; Zou et al., 2011) and is maintained by several quiescence-inducing external stimuli (Scandura et al., 2004; Qian et al., 2007; Blank and Karlsson, 2015).

The presence of cycling stimulation regulates attractor 3. Here, proliferative components such as cyclins are active, while cell cycle inhibitors are inactive. Thus, a proliferative phenotype can be assumed (Orford and Scadden, 2008; Pietras et al., 2011).



		LT-HSC	ST-HSC	Cycling HSC	Unstimulated
	External quiescence	🟩	🟩	🟦	🟦
	External cycling	🟦	🟩	🟩	🟦
Metabolism	PI3K	🟦	🟩	🟩	🟦
	TSC1/2	🟩	🟦	🟦	🟩
	mTORC1	🟦	🟩	🟩	🟦
	FOXO3A	🟩	🟦	🟦	🟦
	ATM	🟩	🟦	🟦	🟦
	ROS	🟦	🟩	🟩	🟩
	Mitochondria	🟦	🟩	🟩	🟦
	Autophagy	🟦	🟦	🟦	🟦
Cell Cycle	RAS	🟦	🟩	🟩	🟦
	ETS	🟦	🟦	🟦	🟦
	MEF	🟦	🟩	🟩	🟦
	GSK3β	🟩	🟦	🟦	🟩
	CTNNB1	🟦	🟩	🟩	🟦
	cMYC	🟦	🟩	🟩	🟦
	BMI1	🟩	🟩	🟩	🟦
	MDM2	🟦	🟩	🟩	🟦
	TP53	🟩	🟦	🟦	🟩
	CDKN1C	🟩	🟩	🟦	🟦
	CDKN1A	🟩	🟦	🟦	🟩
	CDKN1B	🟩	🟦	🟦	🟦
	GFI1	🟩	🟦	🟦	🟩
	RB	🟩	🟩	🟦	🟩
	E2F	🟦	🟦	🟩	🟦
	CCND1	🟦	🟦	🟩	🟦
	CCNE1	🟦	🟦	🟩	🟦
	S-phase	🟦	🟦	🟩	🟦
Survival	AKT	🟦	🟩	🟩	🟦
	CDKN2D	🟦	🟦	🟦	🟩
	CDKN2A	🟦	🟦	🟦	🟩
	Pro-apoptic proteins	🟦	🟦	🟦	🟩
	Anti-apoptotic proteins	🟦	🟩	🟩	🟦
	CYCS	🟦	🟦	🟦	🟩
	Apoptosis	🟦	🟦	🟦	🟩
	Senescence	🟦	🟦	🟦	🟩

FIGURE 2 | Impact of external stimulation. Simulation of the HSC model revealed four attractors that are dependent on the combinations of the two external inputs (external quiescence an external cycling). Components of the model are listed on the left separated into assigned pathways. The state of each component is depicted by colored rectangles. Green indicates active nodes, blue indicates inactive ones.

To sum up, our model is able to recapitulate the three main phenotypes observed for HSC, which we summarized again in **Table 2**. A version of **Table 2** with references for each phenotype describes is provided in the **Supplementary Section 1.2**. Moreover, we are able to show that intrinsic stem cell behavior is affected by external stimulation. This indicates that complex niche-stem cell interactions determine the overall long-term behavior of HSCs.

TABLE 2 | Summary of analyzed attractors.

Attractor	Process	Phenotypical description	Associated HSC phenotype
LT-HCS	Metabolism	Inactive ROS	Quiescent
	Metabolism	Inactive mTORC1	
	Metabolism	Active FOXO3A	
	Cell cycle	Inactive MYC	
	Cell cycle	Active TP53	
	Cell cycle	Active CDKN1C	
	Cell cycle	Active CDKN1A	
	Cell cycle	Active CDKN1B	
	Cell cycle	Active GFI1	
	Cell cycle	Inactive S-phase	
ST-HSC	Metabolism	Active ROS	Activated HSC
	Metabolism	Active mTORC1	
	Metabolism	Inactive FOXO3A	
	Cell cycle	Active MYC	
	Cell cycle	Inactive TP53	
	Cell cycle	Active CDKN1C	
	Cell cycle	Inactive S-phase	
Cycling HSC	Cell cycle	Active CCND1	Proliferating HSC
	Cell cycle	Active CCNE1	
	Cell cycle	Active S-phase	
Unstimulated	External Stimuli	–	–

Main nodes discussed in the text, final biological interpretation and references are reported for each obtained attractor. A version with references is provided in the ***Supplementary Section 1.2****.*

3.2. Awakening of the HSC: Progression From Dormancy to Cycling

Above, we showed the presence of different individual phenotypes in the HSC model. Furthermore, we believe that distinct phenotypes depend on each other. To support this assumption, we investigated whether the single HSC entities transit into each other by alterations of external niche factors (**Figure 3**). This hypothesis is in line with previous findings indicating that LT-HSCs are activated and transit to ST-HSCs which further activated proliferation (Wilson and Trumpp, 2006; Wilson et al., 2007, 2009; Foudi et al., 2009; Pietras et al., 2011). Since we already excluded the possibility of no external stimulation, there are three remaining influences from the niche: only quiescent stimuli, only proliferative (cycling) stimuli, or both stimuli are present.

Hence, we started our simulation by the LT-HSC attractor. We used this attractor state and further activated the external cycling input. This progression leads to a single state attractor describing the ST-HSC phenotype (**Figure 3**). Finally, form the obtained attractor pattern, we further inactivated the external quiescence node and simulated the progression toward the attractor again.

Our simulation of the awakening process started from the LT-HSC matched attractor. This attractor is characterized by active cell cycle inhibitors (CDKN1A, CDKN1B, CDKN1C, GFI1) and inactive RAS/PI3K pathways. The impact of both external stimuli was investigated. The progression toward the attractor is takes ten-time steps (**Figure 3**). First, the presence of external quiescent stimuli leads to up-regulation of cell cycle inhibitors (CDKN1C, CDKN1A, CDKN1B). However, the additional presence of the external cycling stimuli leads to a stabilization of RAS and PI3K signaling (time step zero to four). Furthermore, the stabilization of RAS and AKT induces the activation of cMYC by downregulation of GSK3β. In turn, cMYC leads to loss of cell cycle inhibitor CDKN1A (time step three to seven).

On the other hand, AKT activity induces mTORC1 signaling and represses FOXO3A (time steps three to seven). Thus, activation of mTORC1 directly activates mitochondria and the production of ROS. ROS activation is further sustained by the inhibition of FOXO3A and, consequently, ATM (time steps three to seven). Loss of ATM in the absence of cyclin-dependent kinase inhibitor 2D (CDKN2D) stabilizes MDM2 that, in turn, downregulates TP53 and its downstream target GFI1 (time steps six to ten). Stabilization of active MDM2 is further maintained by activation of myeloid elf-1-like factor (MEF) in response to RAS activation (from time step two). The finally reached attractor is the one that was previously connected to an ST-HSC. This cascade, including both external niche factors, indicates a destabilization of TP53 based on the activity of RAS/AKT and their downstream targets. Finally, starting from the ST-HSC attractor, we simulated our system by removing external quiescence stimuli. This time the system reaches a cycling HSC phenotype in five-time steps. Due to lack of external quiescence input, the phenotype switches toward further evolution of states leading to entry in the cell cycle. In fact, the lack of active cell cycle inhibitors leads to regulation of Retinoblastoma protein (RB), Cyclin D (CCND1) and Cyclin E (CCNE1), and finally to activation of the S-phase entry (time steps one to four) and thus proliferation.

Further, we considered the possibility that a highly quiescent HSC can be directly driven to enter the cell cycle. Hence we considered activation of external cycling stimulation starting from the LT-HSC attractor. The progression toward the cycling attractor presents a similar cascade, as shown in **Figure 3**, and is reported and briefly discussed in the **Supplementary Section 2**.

Summarizing these simulations, our model is not only able to recapitulate the different phenotypes of an HSC (**Figure 2**) but also shows that they depend on each other (**Figure 3** and **Supplementary Section 2**). This indicates a dependency of HSC behavior from niche factors. However, we do not want to underestimate the impact of intrinsic rewiring regulations. The switching cascade between HSC entities that mTORC1 signaling and ROS regulation play a significant role in reasserting the final phenotypes. Besides, we observed that different mechanisms could govern the quiescence status of HSC, making them more or less reactive to external stimuli (**Figure 3**). Starting from an ST-HSC phenotype, the progression toward a cycling HSC happens in fewer time steps than considering an LT-HSC (**Figure 3** and **Supplementary Section 2**). The results of our progressions are in accordance with the fact that ST-HSCs are thought to be more active HSCs: they preserve a quiescent state but can enter the cell cycle (Wilson et al., 2004, 2007, 2009; Foudi et al., 2009).

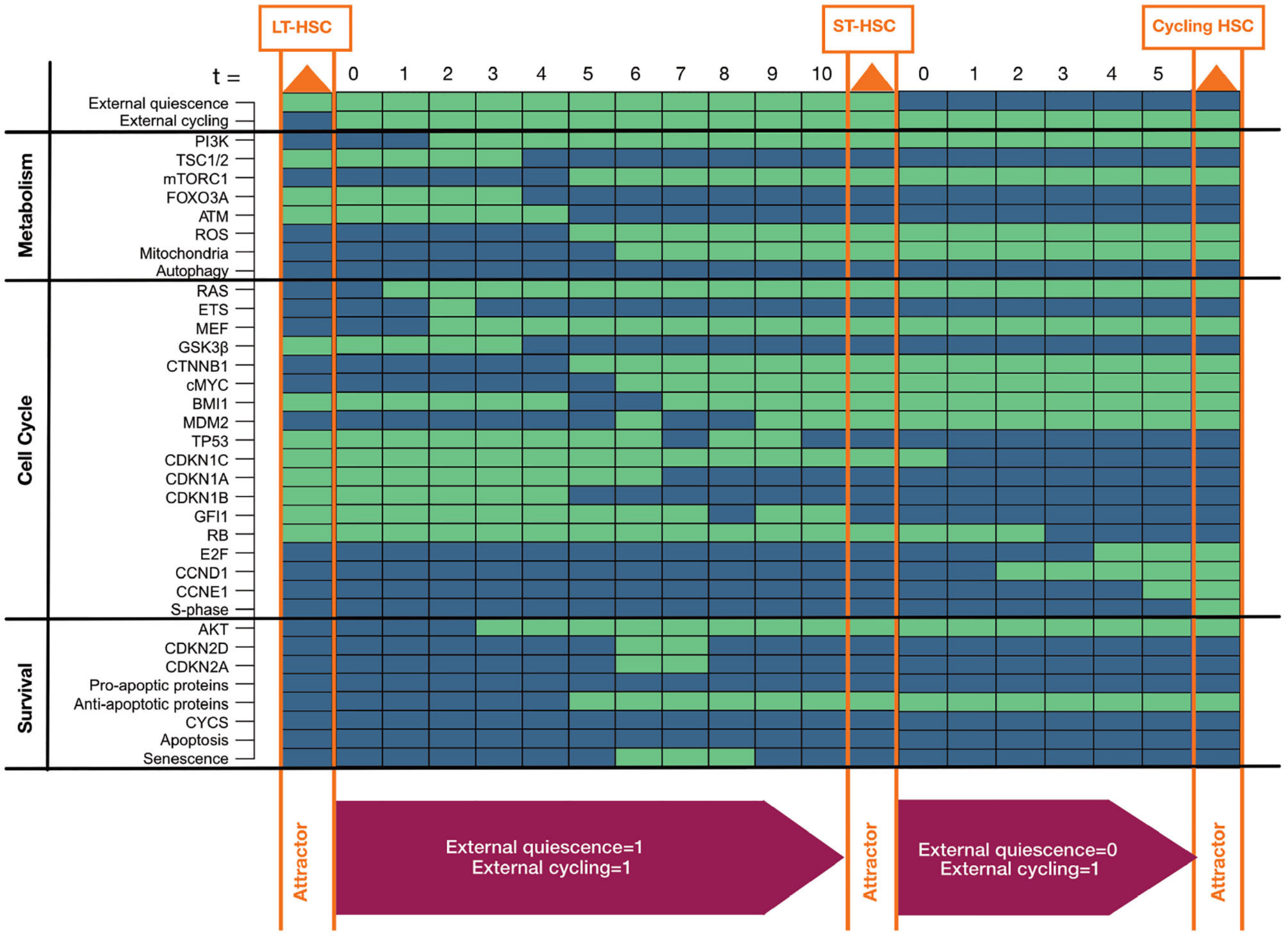


FIGURE 3 | Progression toward awakening. The figure shows progression from LT to ST to cycling HSC depending on activation of external stimuli. Attractors are highlighted in orange in the progression. Activated external stimuli are depicted in magenta boxes below. Components of the model are listed on the left separated into assigned pathways. The state of each component is depicted by colored rectangles. Green indicates active components, blue indicates inactive ones.

3.3. Balancing Quiescence: A Novel Mechanism for TP53 Regulation in HSC

TP53 activity is known to be downregulated by MDM2. Furthermore, MDM2 expression is triggered by TP53 itself, setting a negative feedback loop in its own regulation (Jones et al., 1995; Montes de Oca Luna et al., 1995; Ringshausen et al., 2006). However, we assume that this general mechanism is somehow uneven in HSCs to allow TP53 activity and maintenance of quiescence.

In the previous section, we showed progression toward different HSC phenotypes (**Figure 3** and **Supplementary Section 2**). In our simulation we observed alterations in the TP53 regulation during the cascades leading to the different attractors. In fact, in LT-HSC the activity of TP53 is maintained through a FOXO3A-ATM axis that destabilizes MDM2 activity. In ST-HSC, instead, external cycling signals cause an up regulation of the RAS/PI3K axis. This activation has two major effects that down regulate TP53. First, AKT activation causes a down regulation of FOXO3A and ATM (**Figure 3**, time steps three to seven). Second, RAS activation upregulates MEF, a transcriptional activator of MDM2. As a result, MDM2 and TP53 as well as its downstream target the cell cycle inhibitor GFI1 are downregulated. Hence, we believe that TP53 activity in steady state conditions is balanced by a crosstalk of intrinsic and extrinsic factors, finely tuning its activity in regulating HSC quiescence. This general regulatory hypothesis will be further discussed in the following section.

3.4. *In silico* Interventions on the Model

The hypothesized mechanism of TP53 regulation was further investigated by performing *in silico* knockouts. In the following, we compare our interventions to phenotypes of mouse models. To verify our model's prediction power, we assessed general long-term behavior and its impact on TP53.

3.4.1. FOXO3A and ATM Knockouts: Guardians of ROS Regulation

FOXO3A and ATM are two significant regulators of ROS in quiescent HSCs (Bakker and Passegué, 2013; Ludin et al., 2014). Mouse models for FOXO3A knockout (Tothova and Gilliland,

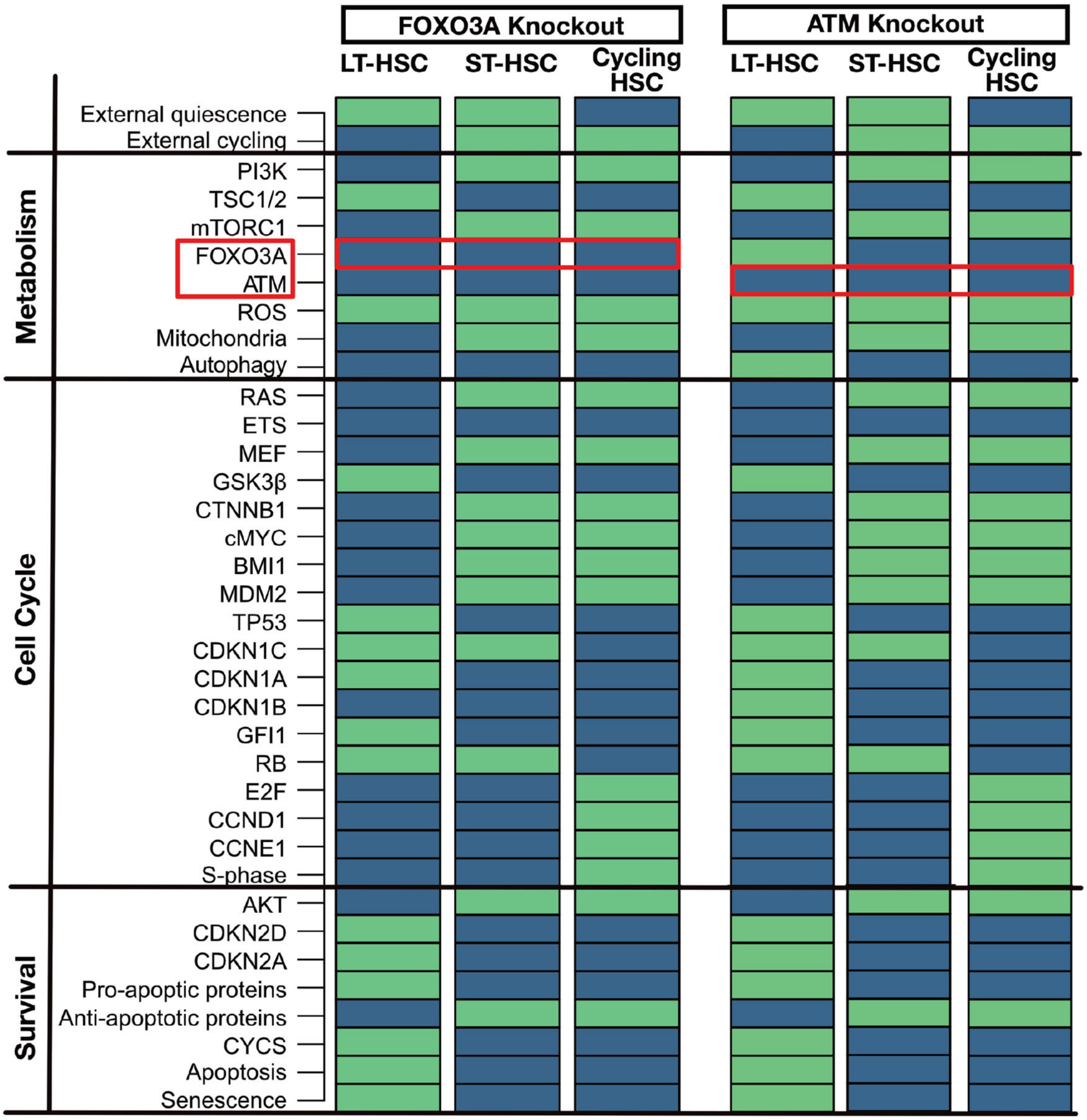


FIGURE 4 | Similar phenotypes of ATM and FOXO3A knockout. An *in silico* knockout of FOXO3A or ATM increases ROS levels but only induces apoptosis and senescence in the LT-HSC population. Proteins are listed on the left with their pathway affiliations. The state of each protein is represented by colored rectangles. Activity is represented in light green and inhibition in dark blue. For simplicity due to the highly similar phenotype of the two *in silico* knockouts only FOXO3A is reported.

2007; Tothova et al., 2007; Yalcin et al., 2008; Maryanovich et al., 2012) and ATM knockout (Ito et al., 2004, 2006; Maryanovich et al., 2012) show increased ROS, CDKN2D, and CDKN2A levels in the HSC compartment leading to senescence and impaired survival. This leads impaired self-renewal and repopulation ability of the LT-HSCs. However, it is still possible to observe a normal differentiation in hematopoiesis (Ito et al., 2004, 2006; Tothova and Gilliland, 2007; Tothova et al., 2007; Yalcin et al., 2008; Maryanovich et al., 2012). The observed phenotypes within the mouse models could be matched to our attractors (**Figure 4**). LT-HSC shows senescence and apoptosis activation and increased ROS levels due to the activation of CDKN2A and CDKN2D. In particular, activation of CDKN2D sustains TP53 activation. The latter, in the presence of ROS, can trigger apoptosis and senescence. ST-HSCs are less affected and do not undergo apoptosis due to the activation of anti-apoptotic proteins and already inactive FOXO3A and ATM. Also, the cycling HSC attractor is not altered. This result is also in line with our cascades showed in **Figure 3**. There it is shown that FOXO3A and ATM are already inactivated in ST-HSC after activation of PI3K/AKT signaling. To summarize, in the FOXO3A and ATM knockouts, our attractors show significant perturbations in the LT-HSC population. Here, activation of ROS, CDKN2A, CDKN2D leads depletion of the LT-HSC population. However,

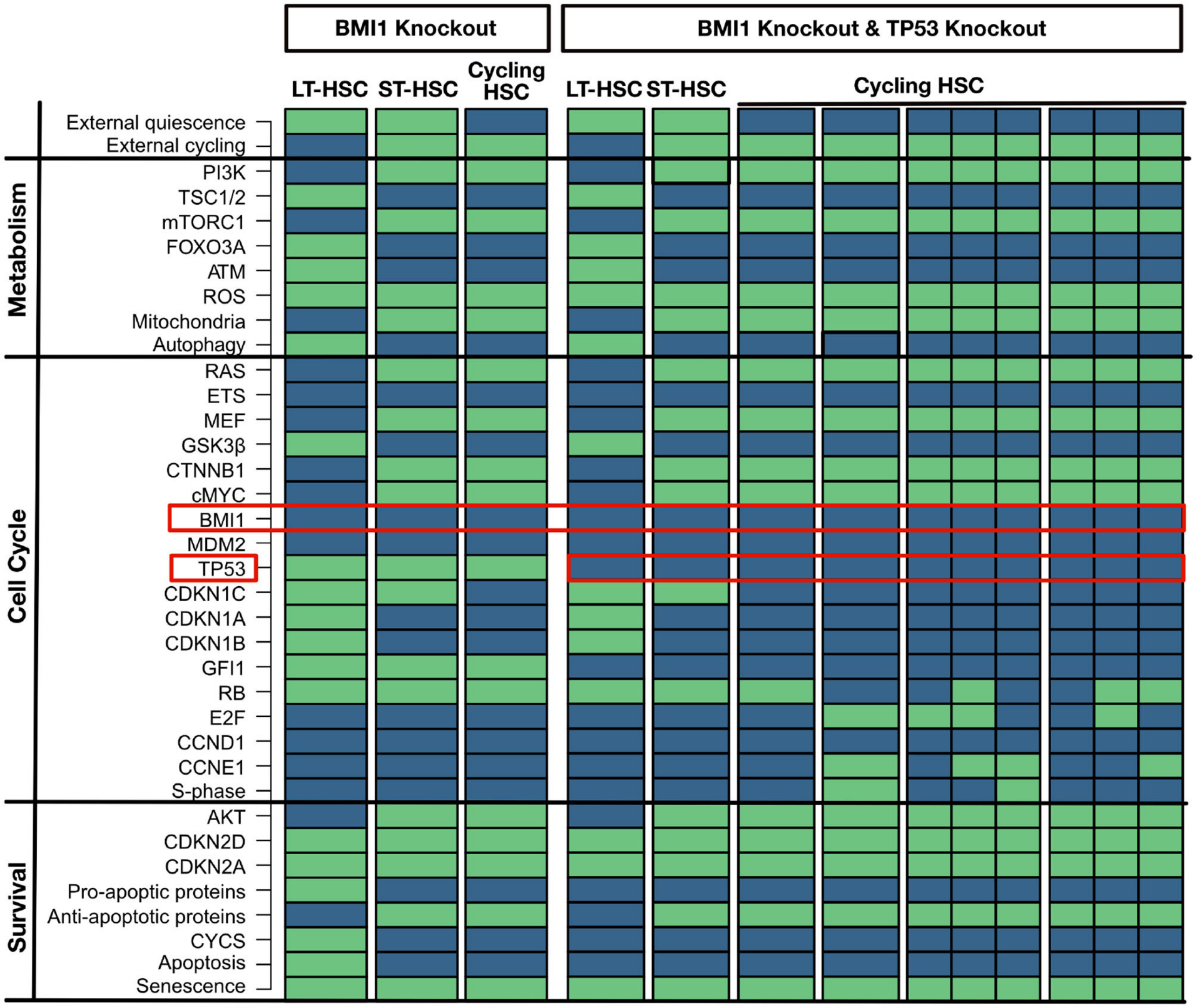


FIGURE 5 | BMI1 loss impairs self-renewal (left part). An *in silico* knockout of BMI1 impairs cell cycle entry for all HSC populations. Proteins are listed on the left with their pathway affiliations. TP53 partially rescue a BMI1 knockout (right part). An *in silico* knockout of BMI1 and TP53 rescues a the impact of a loss of BMI1 by impairing apoptosis. Proteins are listed on the left with their pathway affiliations. The state of each protein is represented by colored rectangles. Activity is represented in light green and inhibition in dark blue.

the other two attractors, namely the one representing ST-HSC and cycling HSC are not affected by the perturbation. This might explain the normal hematopoiesis seen in the FOXO3A and ATM mouse models (Ito et al., 2004; Yalcin et al., 2008). Moreover, in the FOXO3A depleted mice, it was shown that the depletion of the HSCs comes from the LT-HSC population (Yalcin et al., 2008), supporting our population-based results.

3.4.2. BMI1 Knockout: Effect and Rescue of the Gatekeeper of Self-Renewal

BMI1 is a crucial regulator of self-renewal and is widely expressed in the HSC compartment. Therefore, the BMI1 knockout simulation shows a huge effect on the phenotype in mice leading to defect in self-renewal, depletion of the HSC compartment, and shortened life-span (Park et al., 2003; Liu J.et al., 2009; Rizo et al., 2009). This effect is accompanied by increased ROS levels and increased CDKN2A and CDKN2D expression, which finally induces apoptosis and senescence (Park et al., 2003; Liu J.et al., 2009; Rizo et al., 2009). These huge implications are also observable in the attractor pattern. Here, BMI1 loss induces a substantial impairment of cell cycle entry for the HSC in all obtained attractors (**Figure 5**). All attractors show senescence, and the LT-HSC attractor shows apoptosis additionally. The attractor for the cycling HSC does not show activation of S-phase. Based on our cascades shown in **Figure 3**, loss of BMI1 upregulates CDKN2A and CDKN2D, causing impaired quiescence and activation maintenance among the three phenotypes.

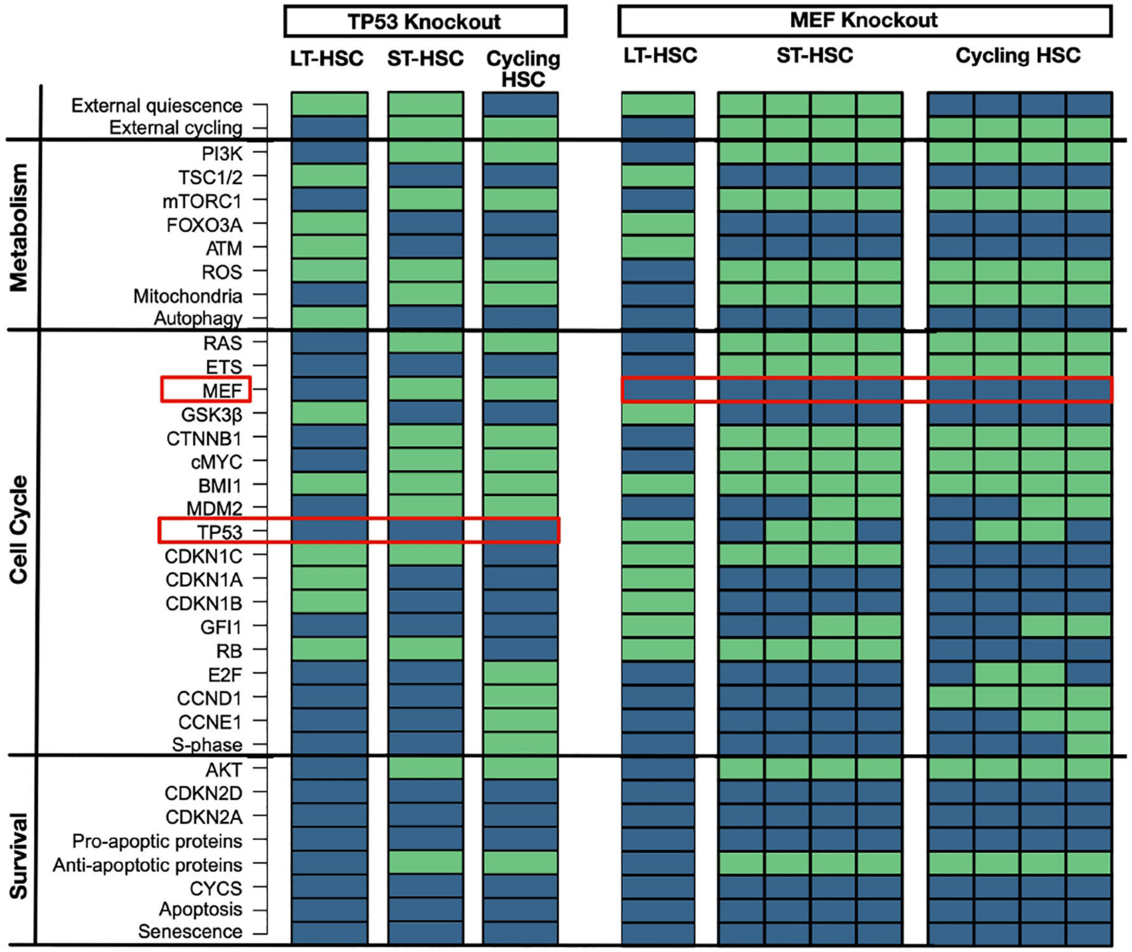


FIGURE 6 | TP53 and MEF knockouts. TP53 knockout has minor impact on HSC populations. An *in silico* knockout of TP53 causes loss of GFI1 in LT-HSCs and increased ROS levels. ST-HSC and cycling ones are not impacted. Proteins are listed on the left with their pathway affiliations. The state of each protein is represented by colored rectangles. Activity is represented in light green and inhibition in dark blue. MEF knockout stabilizes TP53. An *in silico* knockout of MEF stabilized TP53 and enhanced quiescence. Here, ST-HSC and cycling HSCs show activity of GFI1 and cycling HSCs are impaired in entering S-phase. Proteins are listed on the left with their pathway affiliations. The state of each protein is represented by colored rectangles. Activity is represented in light green and inhibition in dark blue.

Various possibilities to partially rescue the phenotype of BMI1 loss have been suggested. One of these suggestions is a loss of TP53 which impairs apoptosis (Akala et al., 2008; Warr et al., 2011). This behavior can also be seen in the attractor when TP53 is perturbed together with BMI1 (**Figure 5**). Again, these simulations show that an increase of ROS and CDKN2D activates TP53 and guide its activity toward pro-apoptotic behavior.

3.4.3. TP53 and MEF Knockouts: Perturbation of Quiescence

TP53 knockout has no major implications on hematopoiesis in mice (TeKippe et al., 2003; Liu Y. et al., 2009; Abbas et al., 2011; Pant et al., 2012). However, loss of GFI1 gives an advantage in competitive assay due to less regulated quiescent state (TeKippe et al., 2003; Liu Y. et al., 2009; Abbas et al., 2011; Pant et al., 2012). A matching attractor pattern is observable in the *in-silico* simulation with no significant induction of apoptosis (**Figure 6**). Moreover, even in our LT-HSC attractor GFI1 gets inactivated. The inactivation of GFI after TP53 was observed in our previous results in the transition between LT and ST-HSC. Hence, our LT-HSC TP53 knockout attractor matches now partially with the ST-HSC (**Figure 2**). Our attractors show also activation of ROS in line with literature results (Liu et al., 2008). Activation of ROS in HSCs has also been connected to a more active state, prone to enter the cell cycle (Bakker and Passegué, 2013; Ludin et al., 2014). Altogether, our results are following experimental observations on TP53 knockout models. After accessing the loss of TP53, we further analyzed the effects of its over stabilization in HSCs. For this reason, we selected another mouse model causing MDM2 downregulation via MEF deletion. MEF knockout mice show a stabilization of TP53 connected to an enhancement of quiescence (Lacorazza et al., 2006; Liu Y. et al., 2009). Likewise, attractor for cycling HSC showed enhanced quiescence due to loss

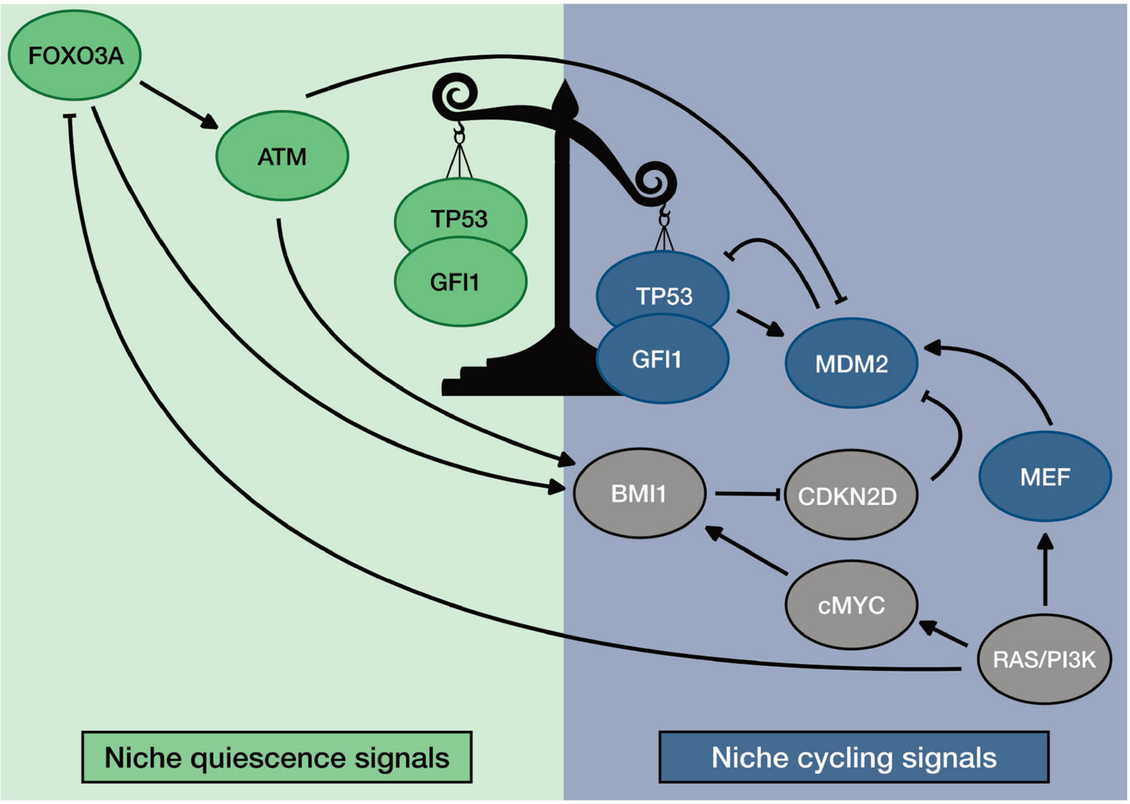


FIGURE 7 | General description of TP53 regulation in HSC. Homeostatic activity of TP53 is promoted by ROS regulators ATM and FOXO3A. Unbalance of TP53 activity is promoted by MEF that stabilizes the transcription of MDM2. Furthermore, the unbalanced activity of TP53 is maintained by niche regulation: quiescence promoting environment favors ROS regulators; Cycling promoting environment promotes MEF expression.

of S-phase. In accordance also the ST-HSC attractor shows now TP53 and GFI1. The stabilization of TP53 affects the transition from one HSC phenotype to another, mainly by the maintenance of GFI1 activity. The unperturbed cascade shows that loss of GFI1 activity happens during the transition between LT- and ST-HSC (**Figure 3**). Hence, in our *in silico* simulations, stabilization of TP53 causes an enhanced quiescent phenotype (**Figure 6**).

Finally, we showed that our model could recapitulate the phenotype of published mouse models. Moreover, perturbation results sustain our hypothesis concerning the general mechanism of regulation of TP53 in HSCs. When ATM, FOXO3A, or BMI1 are depleted, increased ROS and CDKN2D lead to activation of TP53. This effect switches its activity toward the induction of apoptosis and senescence. In homeostatic conditions, instead, they guarantee TP53 activity toward quiescence by downregulating ROS and CDKN2D. Our simulations also showed the relevance of the RAS/PI3K axis in down regulating TP53 when HSCs are activated. The loss of MEF down regulates MDM2 in ST- and cycling HSCs. This loss causes an impaired entry in the S-phase despite the presence of external cycling stimuli. Altogether, our results suggest a general mechanism of regulation of TP53 activity in HSCs. The described mechanism is summarized in **Figure 7**.

3.5. Robustness Analysis

We applied stochastic noise to our Boolean network model to evaluate the robustness—and thus the significance—of our model and the applied simulation. First, the robustness of the network has been computed based on its attractors. Their robustness is measured by the addition of noise by random bit flips in their basin of attraction. We ran the simulation repeatedly, flipping 1, 2, and 3 nodes of each random state. For each number of bit flips, we repeated the simulation three times. Results showed for equal attractors for perturbed and unperturbed states with a mean and standard deviation of 95% (±0.0003)/89% (±0.0001)/85% (±0.0005) for the simulation with 1/2/3 random bit flips per initial node respectively.

As a second measure for robustness, we analyzed the impact of noise on the state transitions in the model and compared it to randomly generated Boolean networks. As before, noise is simulated using random bit flips. We calculate the mean normalized Hamming distance between the successor states of the original state and perturbed copy. Smaller Hamming distance can be interpreted as a better ability to compensate noise. The aim of this analysis is to evaluate the transition robustness of the constructed Boolean network model compared to random networks. Results show a mean Hamming distance of 0.028 for

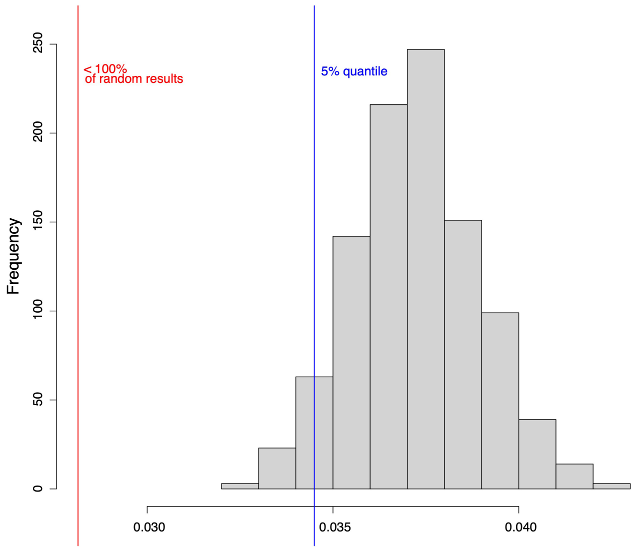


FIGURE 8 | One thousand randomly drawn states of the HSC model were mutated by bit flip and their successor states were computed. The successor states of the mutated and the original states were then compared using the normalized Hamming distance (red line). The same analysis was performed for 1,000 randomly generated networks of the same size (histogram). The blue line shows the 5% quantile.

the constructed HSC model. In contrast, the mean normalized Hamming distance over all random networks is 0.034. The calculated p-value is below 0.0001. Thus, we conclude that our established model is significantly more robust to noise than a set of randomly generated networks of the same size (**Figure 8**).

4. DISCUSSION

In the present work, we proposed an *in silico* Boolean network model depicting the unique wiring that guarantees the maintenance of HSCs in their niche. Our model is able to describe the quiescence mechanism of LT-HSC and ST-HSC. Moreover, we showed the progression toward entry in the cell cycle with respect to changes in external conditions. Additionally, our model strikingly recovered several phenotypes observed in HSCs of knockout mouse models for major intrinsic regulators. This evaluation might be quite challenging under *in vitro* conditions where both isolation and maintenance of HSC require much effort. Identifying methods of isolation and maintenance of HSCs is still challenging (Passegué et al., 2005; Weissman and Shizuru, 2008; Challen et al., 2009; Rossi et al., 2011; Frisch and Calvi, 2014; Jiang et al., 2018; Kobayashi et al., 2019). Hence, the possibility of having an *in silico* mathematical model for simulating the behavior of HSC is of great interest.

External stimuli coming from the niche were summarized as either quiescent or cycling inputs. This choice might be considered an over-simplification of the complex network of signals surrounding the HSC. However, it has been shown that, due to its crucial relevance, HSC maintenance is redundantly regulated by multiple sides. Exemplarily, CDKN1C which is considered the central gatekeeper of quiescence (Passegué et al., 2005; Umemoto et al., 2005; Matsumoto et al., 2011; Tesio and Trumpp, 2011; Zou et al., 2011), is regulated by a multiplicity of endosteal stimuli including TGF-β, TPO and hypoxia (Scandura et al., 2004; Qian et al., 2007; Eliasson et al., 2010; Blank and Karlsson, 2015). Following the idea of redundancy, the loss of hypoxic conditions will not lead to a complete lack of CDKN1C activity (Eliasson et al., 2010). Also, the loss of single developmental pathways has been shown not to affect HSC maintenance, indicating again that a multiplicity of signals conquer the task (Mancini et al., 2005; Maillard et al., 2008; Gao et al., 2009; Hofmann et al., 2009; Kabiri et al., 2015; Oostendorp, 2015). Nevertheless, dissecting too much single influence of each external signal might lead to contradictory results. For example, SCF and TPO have been associated with quiescence *in vivo* (Qian et al., 2007; Yoshihara et al., 2007; Thorén et al., 2008). However, these factors are commonly used to sustain the proliferation of HSC *in vitro* (Ema et al., 2000; Gammaitoni et al., 2003). This differential behavior suggests the possibility that cross-regulation among quiescence signals exists. It has been shown that TGF-β signaling, highly present in the endosteal niche (Yamazaki et al., 2009; Blank and Karlsson, 2015), is able to repress the cytokine-induced activation of PI3K by inhibiting the formation of lipid rafts (Yamazaki et al., 2009). Besides, other mathematical modeling approaches successfully described HSC differentiation by suggesting signaling cumulative environments (Roeder and Lorenz, 2006; Roeder et al., 2008; Glauche et al., 2011). Therefore, it is reasonable to consider an environment with predominant quiescence or cycling stimuli.

Besides single predominant stimuli, we also considered the co-presence of both quiescence and cycling inducing stimuli as sort of gray zones with a balance between the two signals. This consideration is actually in line with different niche models that suggest activated HSCs to be found in regions of the niche that are boundaries between two different environments (Murphy et al., 2005; Wilson and Trumpp, 2006). In this direction Murphy et al., describe HSCs able to self-renew and proliferate to be found in the connection between endosteal and stromal area (Murphy et al., 2005). These HSCs would then sense both quiescence signals coming from the endosteal niche and proliferation stimuli coming from the stroma (Murphy et al., 2005). Also, Trumpp et al. hypothezed a similar regulation, describing the presence of a reservoir quiescent compartment with specialized cells maintaining LT-HSC (Wilson and Trumpp, 2006; Trumpp et al., 2010). Again they characterize a "self-renewing endosteal niche" as the one present at the boundaries of this region when HSCs can get activated and more responsive (Wilson and Trumpp, 2006; Trumpp et al., 2010).

Nevertheless, our results suggest a balance of external signals that sets the instructions for the intrinsic rewiring of the HSC. This means that potentially an LT-HSC could be found in any region having quiescence inducing signals. Thereby expanding the possibility of finding niches for quiescence also in different regions of the bone marrow. This consideration is in accordance to further observations that describe the presence of quiescent HSCs for example in lowly oxygenized regions of sinusoids (Itkin et al., 2016; Pinho and Frenette, 2019).

Different phenotypic characterizations of LT- and ST-HSC describing differential expression patterns have been published (Wilson et al., 2004; Forsberg et al., 2005; Passegué et al., 2005; Jang and Sharkis, 2007). Our attractors could match the primary expression and activity features that identify the two HSC phenotypes. Moreover, we also suggested two different mechanisms of maintaining quiescence in these two subpopulations. Our model indicates that LT-HSC quiescence is preserved by tight regulation of ROS levels from a combination of intrinsic factors induced by external stimuli. It has been observed that keeping oxidative stress low in quiescent HSC is fundamental for preserving their integrity and survival (Jang and Sharkis, 2007; Suda et al., 2011; Ludin et al., 2014). In fact, despite the activity of TP53 and lack of survival stimuli, we do not detect activation of apoptosis due to the highly hampering of cellular stress.

On the other hand, it has been shown that HSCs have a high expression of induced myeloid leukemia cell differentiation protein (MCL1), an anti-apoptotic protein activated in the presence of survival stimuli (Opferman et al., 2005). However, this study was performed on the entire subset of HSC and progenitor cells. Similarly, if we consider our whole set of attractors, we also observe an expression of anti-apoptotic proteins. Sustaining our hypothesis of maintenance of LT-HSC, Kobayashi et al. recently published a study describing the ideal setting for *in vitro* maintenance of quiescent HSCs (Kobayashi et al., 2019). In their work, the authors actually show that hypoxic condition in the presence of residual cytokine stimuli sets the best environment for the maintenance of quiescence (Kobayashi et al., 2019). Our dynamic simulations also lead to a single state attractor depicting a cycling HSC. Initially, this result seems in contradiction with previously published mammalian cell cycle models (Fauré et al., 2006; Tanaka et al., 2017; Diop et al., 2019). In these studies, proliferating cells are represented by cycling attractors. However, we were interested in modeling exit from Gap phase zero (G0) and entering the S-phase. Hence, we selected cell cycle regulators involved in this switch. If we consider only this regulators, also previously published cell cycle models would show a single state attractor (Fauré et al., 2006; Tanaka et al., 2017; Diop et al., 2019).

Moreover, we inserted our starting states of cell cycle regulators from our cycling HSC attractor in the Fauré et al. model and obtained a cyclic attractor (see **Supplementary Section 3**). Altogether, this underlines the correctness of our approach and the extendability of our network model.

Notably, we could also show that LT-HSC and ST-HSC have different responsiveness in terms of time to enter the cell cycle once stimulated. This is also in line with the observation that actually ST-HSC is more reactive and can efficiently enter the cell cycle (Forsberg et al., 2005; Murphy et al., 2005; Wilson et al., 2007, 2008). Our hypothesis or regulatory wiring is further encouraged given the predictive power that our model has toward a variety of *in silico* knockouts that can faithfully recapitulate the phenotype of analog mouse models.

Furthermore, based on our simulation of the progression from a quiescent cell toward the entry in the cycle, we can suggest a mechanism for TP53 homeostatic regulation in HSCs. Here, we propose that the awakening of the HSC follows the initial activation of cellular metabolism. This will then induce the rewiring of the intrinsic network of ROS sensors (FOXO3A and ATM) and finally lead to a loss of cell cycle regulators connected to quiescence such as TP53, CDKN1A, and CDKN1B. In the end, with loss of quiescence signals, the HSC can rapidly enter the cell cycle. Stress sensors have been shown to be fundamental for regulating HSC quiescence (Suda et al., 2011; Bakker and Passegué, 2013; Ludin et al., 2014). The expression of FOXO3A is directly connected to the hypoxic environment, and this leads in turn to the upregulation of active ATM (Yalcin et al., 2008; Suda et al., 2011). The last, besides being an independent regulator of cellular stress, is also involved in the activation of TP53 (Yalcin et al., 2008; Suda et al., 2011; Bakker and Passegué, 2013; Ludin et al., 2014). Thus, our simulations indicate that in LT-HSC ATM is responsible for the regulation of active TP53. This activation is actually new to the HSC maintenance context. This is further enforced from the fact that ATM seems to be required for TP53 immediate regulation of irradiation in quiescent HSC (Mohrin et al., 2010; Kosan and Godman, 2016). Besides its well-defined role in DNA damage response, ATM has already been shown to regulate activation of TP53 under physiological conditions in the context of homeostatic cellular metabolism outside of the HSC context (Armata et al., 2010). Here, It has been suggested that activation of TP53 by ATM is involved in the reduction of ROS (Armata et al., 2010). When instead RAS and PI3K are activated by external stimuli, FOXO3A is downregulated by AKT/mTORC1 signaling potentially also downregulating ATM. This regulation is in accordance with the fact that already in ST-HSC, FOXO3A is excluded from the nucleus (Yalcin et al., 2008). Consequently, in this condition, TP53 is mainly regulated by the MEF/MDM2 and the BMI1/CDKN2D axis (Liu Y. et al., 2009; Abbas et al., 2010; Warr et al., 2011). These regulators set the downregulation of TP53 and GFI1 in the progression toward the ST-HSC phenotype. In accordance, TP53 is downregulated in ST-HSCs (Forsberg et al., 2005; Jang and Sharkis, 2007; Pant et al., 2012). Nevertheless, our model also maintains the hypothesis that TP53 activity is context-dependent (Asai et al., 2011) as shown by our knockout simulations. To further analyze the balanced regulation of TP53 under homeostatic conditions, we performed loss of interaction mutation simulations (see **Supplementary Section 4**). Here, we could show the loss of regulatory axes causing disruption of the original wildtype attractors, when removing the interaction between ATM and MDM2, MEF, and MDM2 or both. In addition, constitutive FOXO3A/ATM or RAS/PI3K in absence of external cycling signals can recapitulate the TP53 activation status (see **Supplementary Section 4**).

Our model of progression is also in accordance with the idea that mTORC1 signaling triggers cellular metabolism and prompts the HSC to enter the cell cycle (Chen et al., 2008, 2009; Rodgers et al., 2014; Baumgartner et al., 2018). Hence, we provide an overall mechanism of HSC awakening and a new general suggestion of TP53 regulation in homeostatic conditions. To our best knowledge, no previous models could show neither

HSC maintenance nor a general mechanistic explanation of TP53 regulation able to summarize observed phenotypes in different experimental conditions.

In addition to our dynamic simulations, we also investigated the robustness of our established Boolean network. Biological networks are considered to be robust toward perturbations. This means that they can adapt to environmental changes, and their functions are resistant to random noise (Kitano, 2002; Greenbury et al., 2010; Graudenzi et al., 2011; Barabási, 2016; Schwab et al., 2020). Hence, we also evaluated the robustness of our network against noise. We both investigated the shift of attractor basin after perturbation and the normalized Hamming distance after bit flip perturbation. In the latter, we compared the performance of our models to one of randomly generated networks of the same size. Equivalent approaches to the ones we presented have been already applied to investigate the robustness of Boolean networks (Joo et al., 2018; Siegle et al., 2018). Altogether, we could show that our model is robust toward random noise.

In conclusion, we established a model able to faithfully recapitulate the regulation of HSC maintenance in the presence of external niche stimuli. Also, we simulated *in silico* the process of awakening of the HSC, giving a mechanistic and overall picture of the process. Further, we suggest a new general regulatory wiring responsible for the regulation of TP53 activity in HSCs. Moreover, we further validated our model by testing its predictive power in a variety of mouse models.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**. All data are deposited in github.

AUTHOR CONTRIBUTIONS

HK, SK, and JS devised the project. NI and SK established the model. NI, SK, and JS performed simulations and were responsible for the visualization. All authors discussed the results and contributed to the final manuscript. HK provided funding and supervised the project.

FUNDING

HK acknowledges funding from the German Federal Ministery of Education and Research (BMBF) e:MED confirm (id 01ZX1708C). Furthermore, HK acknowledges funding from the German Science Foundation [DFG, 217328187 (SFB 1074) and 288342734 (GRK HEIST)].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00848/full#supplementary-material

REFERENCES

Abbas, H. A., Maccio, D. R., Coskun, S., Jackson, J. G., Hazen, A. L., Sills, T. M., et al. (2010). Mdm2 is required for survival of hematopoietic stem cells/progenitors via dampening of ROS-induced p53 activity. *Cell Stem Cell* 7, 606–617. doi: 10.1016/j.stem.2010.09.013

Abbas, H. A., Pant, V., and Lozano, G. (2011). The ups and downs of p53 regulation in hematopoietic stem cells. *Cell Cycle* 10, 3257–3262. doi: 10.4161/cc.10.19.17721

Adolfsson, J., Borge, O. J., Bryder, D., Theilgaard-Mönch, K., Åstrand-Grundström, I., Sitnicka, E., et al. (2001). Upregulation of Flt3 expression within the bone marrow lin?Sca1+c-kit+ stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* 15, 659–669. doi: 10.1016/S1074-7613(01)00220-5

Akala, O. O., Park, I.-K., Qian, D., Pihalja, M., Becker, M. W., and Clarke, M. F. (2008). Long-term haematopoietic reconstitution by $Trp53^{-/-}p16^{Ink4a-/-}p19^{Arf-/-}$ multipotent progenitors. *Nature* 453, 228–232. doi: 10.1038/nature06869

Aldana, M., and Cluzel, P. (2003). A natural class of robust networks. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8710–8714. doi: 10.1073/pnas.1536783100

Armata, H. L., Golebiowski, D., Jung, D. Y., Ko, H. J., Kim, J. K., and Sluss, H. K. (2010). Requirement of the atm/p53 tumor suppressor pathway for glucose homeostasis. *Mol. Cell. Biol.* 30, 5787–5794. doi: 10.1128/MCB.00347-10

Asai, T., Liu, Y., Bae, N., and Nimer, S. D. (2011). The p53 tumor suppressor protein regulates hematopoietic stem cell fate. *J. Cell. Physiol.* 226, 2215–2221. doi: 10.1002/jcp.22561

Athanasiadis, E. I., Botthof, J. G., Andres, H., Ferreira, L., Lio, P., and Cvejic, A. (2017). Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat. Commun.* 8, 1–11. doi: 10.1038/s41467-017-02305-6

Bakker, S. T., and Passegué, E. (2013). Resilient and resourceful: genome maintenance strategies in hematopoietic stem cells. *Exp. Hematol.* 41, 915–923. doi: 10.1016/j.exphem.2013.09.007

Barabási, A.-L. (2016). *Network Science*. Cambridge, MA: Cambridge University Press.

Baumgartner, C., Toifl, S., Farlik, M., Halbritter, F., Scheicher, R., Fischer, I., et al. (2018). An erk-dependent feedback mechanism prevents hematopoietic stem cell exhaustion. *Cell Stem Cell* 22, 879–892. doi: 10.1016/j.stem.2018.05.003

Blank, U., and Karlsson, S. (2015). TGF-β signaling in the control of hematopoietic stem cells. *Blood* 125, 3542–3550. doi: 10.1182/blood-2014-12-618090

Bradford, G. B., Williams, B., Rossi, R., and Bertoncello, I. (1997). Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. *Exp. Hematol.* 25, 445–453.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., et al. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells *in vivo*. *Nature* 518, 542–546. doi: 10.1038/nature14242

Cabezas-Wallscheid, N., Buettner, F., Sommerkamp, P., Klimmeck, D., Ladel, L., Thalheimer, F. B., et al. (2017). Vitamin a-retinoic acid signaling regulates hematopoietic stem cell dormancy. *Cell* 169, 807–823. doi: 10.1016/j.cell.2017.04.018

Challen, G. A., Boles, N., Lin, K.-Y. K., and Goodell, M. A. (2009). Mouse hematopoietic stem cell identification and analysis. *Cytometry A*, 75A, 14–24. doi: 10.1002/cyto.a.20674

Chen, C., Liu, Y., Liu, R., Ikenoue, T., Guan, K.-L., Liu, Y., et al. (2008). TSC-mTOR maintains quiescence and function of hematopoietic stem cells by repressing mitochondrial biogenesis and reactive oxygen species. *J. Exp. Med.* 205, 2397–2408. doi: 10.1084/jem.20081297

Chen, C., Liu, Y., Liu, Y., and Zheng, P. (2009). The axis of mTOR-mitochondria-ROS and stemness of the hematopoietic stem cells. *Cell Cycle* 8, 1158–1160. doi: 10.4161/cc.8.8.8139

Cheshier, S. H., Morrison, S. J., Liao, X., and Weissman, I. L. (1999). *In vivo* proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3120–3125. doi: 10.1073/pnas.96.6.3120

Coskun, S., and Hirschi, K. K. (2010). Establishment and regulation of the HSC niche: roles of osteoblastic and vascular compartments. *Birth Defects Res*. 90, 229–242. doi: 10.1002/bdrc.20194

Dahlhaus, M., Burkovski, A., Hertwig, F., Müssel, C., Volland, R., Fischer, M., et al. (2016). Boolean modeling identifies Greatwall/MASTL as an important regulator in the AURKA network of neuroblastoma. *Cancer Lett*. 371, 79–89. doi: 10.1016/j.canlet.2015.11.025

Diop, O., Tourniel, L., and Fromion, V. (2019). "Summarizing complex asynchronous Boolean attractors, application to the analysis of a mammalian cell cycle model," in *2019 18th European Control Conference (ECC)*, ed F. Garofolo (Naples), 1677–1682. doi: 10.23919/ECC.2019.8795712

Eliasson, P., Rehn, M., Hammar, P., Larsson, P., Sirenko, O., Flippin, L. A., et al. (2010). Hypoxia mediates low cell-cycle activity and increases the proportion of long-term-reconstituting hematopoietic stem cells during *in vitro* culture. *Exp. Hematol*. 38, 301–310.e2. doi: 10.1016/j.exphem.2010.01.005

Ema, H., Takano, H., Sudo, K., and Nakauchi, H. (2000). *In vitro* self-renewal division of hematopoietic stem cells. *J. Exp. Med*. 192, 1281–1288. doi: 10.1084/jem.192.9.1281

Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22, e124–e131. doi: 10.1093/bioinformatics/btl210

Fleming, W., Alpern, E., Uchida, N., Ikuta, K., Spangrude, G., and Weissman, I. (1993). Functional heterogeneity is associated with the cell cycle status of murine hematopoietic stem cells. *J. Cell Biol*. 122, 897–902. doi: 10.1083/jcb.122.4.897

Forsberg, E. C., Prohaska, S. S., Katzman, S., Heffner, G. C., Stuart, J. M., and Weissman, I. L. (2005). Differential expression of novel potential regulators in hematopoietic stem cells. *PLoS Genet*. 1:e28. doi: 10.1371/journal.pgen.0010028

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J. W., Jaenisch, R., Carey, V., et al. (2009). Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat. Biotechnol*. 27, 84–90. doi: 10.1038/nbt.1517

Frisch, B. J., and Calvi, L. M. (2014). Hematopoietic stem cell cultures and assays. *Methods Mol. Biol*. 1130, 315–324. doi: 10.1007/978-1-62703-989-5_24

Gammaitoni, L., Bruno, S., Sanavio, F., Gunetti, M., Kollet, O., Cavalloni, G., et al. (2003). *Ex vivo* expansion of human adult stem cells capable of primary and secondary hemopoietic reconstitution. *Exp. Hematol*. 31, 261–270. doi: 10.1016/S0301-472X(02)01077-9

Gao, J., Graves, S., Koch, U., Liu, S., Jankovic, V., Buonamici, S., et al. (2009). Hedgehog signaling is dispensable for adult hematopoietic stem cell function. *Cell Stem Cell* 4, 548–558. doi: 10.1016/j.stem.2009.03.015

Geest, C. R., and Coffer, P. J. (2009). MAPK signaling pathways in the regulation of hematopoiesis. *J. Leukocyte Biol*. 86, 237–250. doi: 10.1189/jlb.0209097

Geiger, H., Koehler, A., and Gunzer, M. (2007). Stem cells, aging, niche, adhesion and Cdc42: a model for changes in cell-cell interactions and hematopoietic stem cell aging. *Cell Cycle* 6, 884–887. doi: 10.4161/cc.6.8.4131

Glauche, I., Thielecke, L., and Roeder, I. (2011). Cellular aging leads to functional heterogeneity of hematopoietic stem cells: a modeling perspective. *Aging Cell* 10, 457–465. doi: 10.1111/j.1474-9726.2011.00692.x

Graudenzi, A., Serra, R., Villani, M., Colacci, A., and Kauffman, S. A. (2011). Robustness analysis of a boolean model of gene regulatory network with memory. *J. Comput. Biol*. 18, 559–577. doi: 10.1089/cmb.2010.0224

Greenbury, S. F., Johnston, I. G., Smith, M. A., Doye, J. P., and Louis, A. A. (2010). The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks. *J. Theor. Biol*. 267, 48–61. doi: 10.1016/j.jtbi.2010.08.006

Guerrouahen, B. S., Al-Hijji, I., and Tabrizi, A. R. (2011). Osteoblastic and vascular endothelial niches, their control on normal hematopoietic stem cells, and their consequences on the development of leukemia. *Stem Cells Int*. 2011:375857-8. doi: 10.4061/2011/375857

Hamey, F. K., Nestorowa, S., Kinston, S. J., Kent, D. G., Wilson, N. K., and Göttgens, B. (2017). Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. *Proc. Natl. Acad. Sci. U.S.A*. 114, 5822–5829. doi: 10.1073/pnas.1610609114

Herrmann, F., Groß, A., Zhou, D., Kestler, H. A., and Kühl, M. (2012). A Boolean model of the cardiac gene regulatory network determining first and second heart field identity. *PLoS ONE* 7:e46798. doi: 10.1371/journal.pone.0046798

Ho, T. T., Warr, M. R., Adelman, E. R., Lansinger, O. M., Flach, J., Verovskaya, E. V., et al. (2017). Autophagy maintains the metabolism and function of young and old stem cells. *Nature* 543, 205–210. doi: 10.1038/nature21388

Hofmann, I., Stover, E. H., Cullen, D. E., Mao, J., Morgan, K. J., Lee, B. H., et al. (2009). Hedgehog signaling is dispensable for adult murine hematopoietic stem cell function and hematopoiesis. *Cell Stem Cell* 4, 559–567. doi: 10.1016/j.stem.2009.03.016

Hopfensitz, M., Müssel, C., Maucher, M., and Kestler, H. A. (2013). Attractors in Boolean networks: a tutorial. *Comput. Stat*. 28, 19–36. doi: 10.1007/s00180-012-0324-2

Itkin, T., Gur-Cohen, S., Spencer, J. A., Schajnovitz, A., Ramasamy, S. K., Kusumbe, A. P., et al. (2016). Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* 532, 323–328. doi: 10.1038/nature17624

Ito, K., Hirao, A., Arai, F., Matsuoka, S., Takubo, K., Hamaguchi, I., et al. (2004). Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature* 431, 997–1002. doi: 10.1038/nature02989

Ito, K., Hirao, A., Arai, F., Takubo, K., Matsuoka, S., Miyamoto, K., et al. (2006). Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nat. Med*. 12, 446–451. doi: 10.1038/nm1388

Jagannathan-Bogdan, M., and Zon, L. I. (2013). Hematopoiesis. *Development* 140, 2463–2467. doi: 10.1242/dev.083147

Jang, Y.-Y., and Sharkis, S. J. (2007). A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. *Blood* 110, 3056–3063. doi: 10.1182/blood-2007-05-087759

Jiang, M., Chen, H., Lai, S., Wang, R., Qiu, Y., Ye, F., et al. (2018). Maintenance of human haematopoietic stem and progenitor cells *in vitro* using a chemical cocktail. *Cell Discov*. 4:59. doi: 10.1038/s41421-018-0059-5

Jones, D. L., and Wagers, A. J. (2008). No place like home: anatomy and function of the stem cell niche. *Nat. Rev. Mol. Cell Biol*. 9, 11–21. doi: 10.1038/nrm2319

Jones, S. N., Roe, A. E., Donehower, L. A., and Bradley, A. (1995). Rescue of embryonic lethality in mdm2-deficient mice by absence of p53. *Nature* 378, 206–208. doi: 10.1038/378206a0

Joo, J. I., Zhou, J. X., Huang, S., and Cho, K.-H. (2018). Determining relative dynamic stability of cell states using boolean network model. *Sci. Rep*. 8, 1–14. doi: 10.1038/s41598-018-30544-0

Kabiri, Z., Numata, A., Kawasaki, A., Edison, B., Tenen, D. G., and Virshup, D. M. (2015). Wnts are dispensable for differentiation and self-renewal of adult murine hematopoietic stem cells. *Blood* 126, 1086–1094. doi: 10.1182/blood-2014-09-598540

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol*. 22, 437–467. doi: 10.1016/0022-5193(69)90015-0

Kauffman, S. A. (1993). *The Origins of Order. Self Organization and Selection in Evolution*. New York, NY: Oxford University Press. doi: 10.1007/978-94-015-8054-0_8

Kauffman, S. A., Peterson, C., Samuelsson, B., and Troein, C. (2004). Genetic networks with canalyzing Boolean rules are always stable. *Proc. Natl. Acad. Sci. U.S.A*. 101, 17102–17107. doi: 10.1073/pnas.0407783101

Kaushansky, K. (2006). Lineage-specific hematopoietic growth factors. *N. Engl. J. Med*. 354, 2034–2045. doi: 10.1056/NEJMra052706

Kiel, M. J., Yilmaz, O. H., Iwashita, T., Yilmaz, O. H., Terhorst, C., and Morrison, S. J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* 121, 1109–1121. doi: 10.1016/j.cell.2005.05.026

Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210. doi: 10.1038/nature01254

Kobayashi, H., Morikawa, T., Okinaga, A., Hamano, F., Hashidate-Yoshida, T., Watanuki, S., et al. (2019). Environmental optimization enables maintenance of quiescent hematopoietic stem cells *ex vivo*. *Cell Rep*. 28, 145–158.e9. doi: 10.1016/j.celrep.2019.06.008

Kopp, H.-G., Avecilla, S. T., Hooper, A. T., and Rafii, S. (2005). The bone marrow vascular niche: home of HSC differentiation and mobilization. *Physiology* 20, 349–356. doi: 10.1152/physiol.00025.2005

Kosan, C., and Godman, M. (2016). Genetic and epigenetic mechanisms that maintain hematopoietic stem cell function. *Stem Cells Int*. 2016:5178965. doi: 10.1155/2016/5178965

Krumsiek, J., Marr, C., Schroeder, T., and Theis, F. J. (2011). Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS ONE* 6:e22649. doi: 10.1371/journal.pone.0022649

Lacorazza, H. D., Yamada, T., Liu, Y., Miyata, Y., Sivina, M., Nunes, J., et al. (2006). The transcription factor MEF/ELF4 regulates the quiescence of primitive hematopoietic cells. *Cancer Cell* 9, 175–187. doi: 10.1016/j.ccr.2006.02.017

Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. *Nature* 553, 418–426. doi: 10.1038/nature25022

Lilly, A. J., Johnson, W. E., and Bunce, C. M. (2011). The haematopoietic stem cell niche: new insights into the mechanisms regulating haematopoietic stem cell behaviour. *Stem Cells Int*. 2011, 274564–10. doi: 10.4061/2011/274564

Liu, B., Chen, Y., and Clair, D. K. S. (2008). Ros and p53: a versatile partnership. *Free Radic. Biol. Med*. 44, 1529–1535. doi: 10.1016/j.freeradbiomed.2008.01.011

Liu, J., Cao, L., Chen, J., Song, S., Lee, I. H., Quijano, C., et al. (2009). Bmi1 regulates mitochondrial function and the DNA damage response pathway. *Nature* 459, 387–392. doi: 10.1038/nature08040

Liu, Y., Elf, S. E., Miyata, Y., Sashida, G., Liu, Y., Huang, G., et al. (2009). p53 regulates hematopoietic stem cell quiescence. *Cell Stem Cell* 4, 37–48. doi: 10.1016/j.stem.2008.11.006

Ludin, A., Gur-Cohen, S., Golan, K., Kaufmann, K. B., Itkin, T., Medaglia, C., et al. (2014). Reactive oxygen species regulate hematopoietic stem cell self-renewal, migration and development, as well as their bone marrow microenvironment. *Antioxid. Redox Signal*. 21, 1605–1619. doi: 10.1089/ars.2014.5941

Macaulay, I. C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., et al. (2016). Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in hematopoietic cells. *Cell Rep*. 14, 966–977. doi: 10.1016/j.celrep.2015.12.082

Maillard, I., Koch, U., Dumortier, A., Shestova, O., Xu, L., Sai, H., et al. (2008). Canonical notch signaling is dispensable for the maintenance of adult hematopoietic stem cells. *Cell Stem Cell* 2, 356–366. doi: 10.1016/j.stem.2008.02.011

Mancini, S. J. C., Mantei, N., Dumortier, A., Suter, U., MacDonald, H. R., and Radtke, F. (2005). Jagged1-dependent Notch signaling is dispensable for hematopoietic stem cell self-renewal and differentiation. *Blood* 105, 2340–2342. doi: 10.1182/blood-2004-08-3207

Maryanovich, M., Oberkovitz, G., Niv, H., Vorobiyov, L., Zaltsman, Y., Brenner, O., et al. (2012). The ATM-BID pathway regulates quiescence and survival of haematopoietic stem cells. *Nat. Cell Biol*. 14, 535–541. doi: 10.1038/ncb2468

Matsumoto, A., Takeishi, S., Kanie, T., Susaki, E., Onoyama, I., Tateishi, Y., et al. (2011). p57 is required for quiescence and maintenance of adult hematopoietic stem cells. *Cell Stem Cell* 9, 262–271. doi: 10.1016/j.stem.2011.06.014

Meek, D. W. (2009). Tumour suppression by p53: a role for the DNA damage response? *Nat. Rev. Cancer* 9, 714–723. doi: 10.1038/nrc2716

Meyer, P., Maity, P., Burkovski, A., Schwab, J., Müssel, C., Singh, K., et al. (2017). A model of the onset of the senescence associated secretory phenotype after DNA damage induced senescence. *PLoS Comput. Biol.* 13:e1005741. doi: 10.1371/journal.pcbi.1005741

Mohrin, M., Bourke, E., Alexander, D., Warr, M. R., Barry-Holson, K., Le Beau, M. M., et al. (2010). Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* 7, 174–185. doi: 10.1016/j.stem.2010.06.014

Montes de Oca Luna, R., Wagner, D. S., and Lozano, G. (1995). Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. *Nature* 378, 203–206. doi: 10.1038/378203a0

Murphy, M. J., Wilson, A., and Trumpp, A. (2005). More than just proliferation: Myc function in stem cells. *Trends Cell Biol.* 15, 128–137. doi: 10.1016/j.tcb.2005.01.008

Müssel, C., Hopfensitz, M., and Kestler, H. A. (2010). BoolNet–an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* 26, 1378–1380. doi: 10.1093/bioinformatics/btq124

Olariu, V., and Peterson, C. (2019). Kinetic models of hematopoietic differentiation. *WIREs Syst. Biol. Med*. 11:e1424. doi: 10.1002/wsbm.1424

Oostendorp, R. A. J. (2015). Secretion of Wnts is dispensable for hematopoiesis. *Blood* 126, 1051–1052. doi: 10.1182/blood-2015-07-653402

Opferman, J. T., Iwasaki, H., Ong, C. C., Suh, H., Mizuno, S.-I., Akashi, K., et al. (2005). Obligate role of anti-apoptotic MCL-1 in the survival of hematopoietic stem cells. *Science* 307, 1101–1104. doi: 10.1126/science.1106114

Orford, K. W., and Scadden, D. T. (2008). Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat. Rev. Genet*. 9, 115–128. doi: 10.1038/nrg2269

Pant, V., Quintás-Cardama, A., and Lozano, G. (2012). The p53 pathway in hematopoiesis: lessons from mouse models, implications for humans. *Blood* 120, 5118–5127. doi: 10.1182/blood-2012-05-356014

Park, I.-K., Qian, D., Kiel, M., Becker, M. W., Pihalja, M., Weissman, I. L., et al. (2003). BMI-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* 423, 302–305. doi: 10.1038/nature01587

Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C., and Weissman, I. L. (2005). Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med*. 202, 1599–1611. doi: 10.1084/jem.20050967

Pietras, E. M., Warr, M. R., and Passegué, E. (2011). Cell cycle regulation in hematopoietic stem cells. *J. Cell Biol*. 195, 709–720. doi: 10.1083/jcb.201102131

Pinho, S., and Frenette, P. S. (2019). Haematopoietic stem cell activity and interactions with the niche. *Nat. Rev. Mol. Cell Biol*. 20, 303–320. doi: 10.1038/s41580-019-0103-9

Qian, H., Buza-Vidas, N., Hyland, C. D., Jensen, C. T., Antonchuk, J., Månsson, R., et al. (2007). Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* 1, 671–684. doi: 10.1016/j.stem.2007.10.008

Qu, X., Aldana, M., and Kadanoff, L. P. (2002). Numerical and theoretical studies of noise effects in the Kauffman model. *J. Stat. Phys*. 109, 967–986. doi: 10.1023/A:1020416308456

Ringshausen, I., O'Shea, C. C., Finch, A. J., Swigart, L. B., and Evan, G. I. (2006). MDM2 is critically and continuously required to suppress lethal p53 activity *in vivo. Cancer Cell* 10, 501–514. doi: 10.1016/j.ccr.2006.10.010

Rizo, A., Olthof, S., Han, L., Vellenga, E., de Haan, G., and Schuringa, J. J. (2009). Repression of BMI1 in normal and leukemic human CD34+ cells impairs self-renewal and induces apoptosis. *Blood* 114, 1498–1505. doi: 10.1182/blood-2009-03-209734

Rodgers, J. T., King, K. Y., Brett, J. O., Cromie, M. J., Charville, G. W., Maguire, K. K., et al. (2014). mTORC1 controls the adaptive transition of quiescent stem cells from G_0 to G_{Alert}. *Nature* 510, 393–396. doi: 10.1038/nature13255

Roeder, I., Horn, K., Sieburg, H.-B., Cho, R., Muller-Sieburg, C., and Loeffler, M. (2008). Characterization and quantification of clonal heterogeneity among hematopoietic stem cells: a model-based approach. *Blood* 112, 4874–4883. doi: 10.1182/blood-2008-05-155374

Roeder, I., and Lorenz, R. (2006). Asymmetry of stem cell fate and the potential impact of the niche. *Stem Cell Rev*. 2, 171–180. doi: 10.1007/s12015-006-0045-4

Rossi, L., Challen, G. A., Sirin, O., Lin, K. K.-Y., and Goodell, M. A. (2011). Hematopoietic stem cell characterization and isolation. *Methods Mol. Biol*. 750, 47–59. doi: 10.1007/978-1-61779-145-1_3

Sawai, C. M., Babovic, S., Upadhaya, S., Knapp, D. J. H. F., Lavin, Y., Lau, C. M., et al. (2016). Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. *Immunity* 45, 597–609. doi: 10.1016/j.immuni.2016.08.007

Scandura, J. M., Boccuni, P., Massagué, J., and Nimer, S. D. (2004). Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation. *Proc. Natl. Acad. Sci. U.S.A*. 101, 15231–15236. doi: 10.1073/pnas.0406771101

Schofield, R. (1978). The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* 4, 7–25.

Schwab, J., Burkovski, A., Siegle, L., Müssel, C., and Kestler, H. A. (2017). ViSiBooL-visualization and simulation of Boolean networks with temporal constraints. *Bioinformatics* 33, 601–604. doi: 10.1093/bioinformatics/btw661

Schwab, J. D., Kühlwein, S. D., Ikonomi, N., Kühl, M., and Kestler, H. A. (2020). Concepts in Boolean network modeling: what do they all mean? *Comput. Struct. Biotechnol. J*. 18, 571–582. doi: 10.1016/j.csbj.2020.03.001

Siegle, L., Schwab, J. D., Kühlwein, S. D., Lausser, L., Tümpel, S., Pfister, A. S., et al. (2018). A Boolean network of the crosstalk between IGF and Wnt signaling in aging satellite cells. *PLoS ONE* 13:e0195126. doi: 10.1371/journal.pone.0195126

Spangrude, G., Heimfeld, S., and Weissman, I. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science* 241, 58–62. doi: 10.1126/science.2898810

Suda, T., Takubo, K., and Semenza, G. L. (2011). Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell* 9, 298–310. doi: 10.1016/j.stem.2011.09.010

Sugiyama, T., Kohara, H., Noda, M., and Nagasawa, T. (2006). Maintenance of the hematopoietic stem cell pool by CXCL14-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* 25, 977–988. doi: 10.1016/j.immuni.2006.10.016

Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J., et al. (2014). Clonal dynamics of native haematopoiesis. *Nature* 514, 322–327. doi: 10.1038/nature13824

Tanaka, H., Fauré, A., and Matsuno, H. (2017). "Boolean modeling of mammalian cell cycle and cancer pathways," in *International Conference on Artificial Life and Robotics 2017*, ed M. Sugisaka (Miyazaki), 507–510. doi: 10.5954/ICAROB.2017.GS4-3

TeKippe, M., Harrison, D. E., and Chen, J. (2003). Expansion of hematopoietic stem cell phenotype and activity in Trp53-null mice. *Exp. Hematol.* 31, 521–527. doi: 10.1016/S0301-472X(03)00072-9

Tesio, M., and Trumpp, A. (2011). Breaking the cell cycle of HSCs by p57 and friends. *Cell Stem Cell* 9, 187–192. doi: 10.1016/j.stem.2011.08.005

Thomas, R., and Kaufman, M. (2001). Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos* 11, 180–195. doi: 10.1063/1.1349893

Thorén, L. A., Liuba, K., Bryder, D., Nygren, J. M., Jensen, C. T., Qian, H., et al. (2008). Kit regulates maintenance of quiescent hematopoietic stem cells. *J. Immunol.* 180, 2045–2053. doi: 10.4049/jimmunol.180.4.2045

Toledo, F., and Wahl, G. M. (2006). Regulating the p53 pathway: *in vitro* hypotheses, *in vivo* veritas. *Nat. Rev. Cancer* 6, 909–923. doi: 10.1038/nrc2012

Tothova, Z., and Gilliland, D. G. (2007). FoxO transcription factors and stem cell homeostasis: insights from the hematopoietic system. *Cell Stem Cell* 1, 140–152. doi: 10.1016/j.stem.2007.07.017

Tothova, Z., Kollipara, R., Huntly, B. J., Lee, B. H., Castrillon, D. H., Cullen, D. E., et al. (2007). FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* 128, 325–339. doi: 10.1016/j.cell.2007.01.003

Trumpp, A., Essers, M., and Wilson, A. (2010). Awakening dormant haematopoietic stem cells. *Nat. Rev. Immunol.* 10, 201–209. doi: 10.1038/nri2726

Umemoto, T., Yamato, M., Nishida, K., Yang, J., Tano, Y., and Okano, T. (2005). p57Kip2 is expressed in quiescent mouse bone marrow side population cells. *Biochem. Biophys. Res. Commun.* 337, 14–21. doi: 10.1016/j.bbrc.2005.09.008

van Os, R., de Haan, G., and Dykstra, B. J. (2009). Hematopoietic stem cell quiescence: yet another role for p53. *Cell Stem Cell* 4, 7–8. doi: 10.1016/j.stem.2008.12.007

Vousden, K. H., and Lane, D. P. (2007). p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* 8, 275–283. doi: 10.1038/nrm2147

Wagner, W., Horn, P., Bork, S., and Ho, A. D. (2008). Aging of hematopoietic stem cells is regulated by the stem cell niche. *Exp. Gerontol.* 43, 974–980. doi: 10.1016/j.exger.2008.04.007

Warr, M. R., Pietras, E. M., and Passegué, E. (2011). Mechanisms controlling hematopoietic stem cell functions during normal hematopoiesis and hematological malignancies. *WIREs Syst. Biol. Med.* 3, 681–701. doi: 10.1002/wsbm.145

Weissman, I. L., and Shizuru, J. A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. *Blood* 112, 3543–3553. doi: 10.1182/blood-2008-08-078220

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., et al. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* 135, 1118–1129. doi: 10.1016/j.cell.2008.10.048

Wilson, A., Laurenti, E., and Trumpp, A. (2009). Balancing dormant and self-renewing hematopoietic stem cells. *Curr. Opin. Genet. Dev.* 19, 461–468. doi: 10.1016/j.gde.2009.08.005

Wilson, A., Murphy, M. J., Oskarsson, T., Kaloulis, K., Bettess, M. D., Oser, G. M., et al. (2004). c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* 18, 2747–2763. doi: 10.1101/gad.313104

Wilson, A., Oser, G. M., Jaworski, M., Blanco-Bose, W. E., Laurenti, E., Adolphe, C., et al. (2007). Dormant and self-renewing hematopoietic stem cells and their niches. *Ann. N. Y. Acad. Sci.* 1106, 64–75. doi: 10.1196/annals.1392.021

Wilson, A., and Trumpp, A. (2006). Bone-marrow haematopoietic-stem-cell niches. *Nat. Rev. Immunol.* 6, 93–106. doi: 10.1038/nri1779

Yalcin, S., Zhang, X., Luciano, J. P., Mungamuri, S. K., Marinkovic, D., Vercherat, C., et al. (2008). Foxo3 is essential for the regulation of ataxia telangiectasia mutated and oxidative stress-mediated homeostasis of hematopoietic stem cells. *J. Biol. Chem.* 283, 25692–25705. doi: 10.1074/jbc.M800517200

Yamazaki, S., Iwama, A., Takayanagi, S.-I., Eto, K., Ema, H., and Nakauchi, H. (2009). TGF-beta as a candidate bone marrow niche signal to induce hematopoietic stem cell hibernation. *Blood* 113, 1250–1256. doi: 10.1182/blood-2008-04-146480

Yokota, T. (2019). "Hierarchy" and "holacracy" a paradigm of the hematopoietic system. *Cells* 8:1138. doi: 10.3390/cells8101138

Yoshihara, H., Arai, F., Hosokawa, K., Hagiwara, T., Takubo, K., Nakamura, Y., et al. (2007). Thrombopoietin/MPL signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. *Cell Stem Cell* 1, 685–697. doi: 10.1016/j.stem.2007.10.020

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.-G., et al. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* 425, 836–841. doi: 10.1038/nature02041

Zhang, Y., Gao, S., Xia, J., and Liu, F. (2018). Hematopoietic hierarchy-an updated roadmap. *Trends Cell Biol.* 28, 976–986. doi: 10.1016/j.tcb.2018.06.001

Zou, P., Yoshihara, H., Hosokawa, K., Tai, I., Shinmyozu, K., Tsukahara, F., et al. (2011). p57(Kip2) and p27(Kip1) cooperate to maintain hematopoietic stem cell quiescence through interactions with Hsc70. *Cell Stem Cell* 9, 247–261. doi: 10.1016/j.stem.2011.07.003

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 12 August 2020
doi: 10.3389/fphys.2020.01012

Extracting Insights From Temporal Data by Integrating Dynamic Modeling and Machine Learning

*Richard Ballweg†, Kristen A. Engevik†, Marshall H. Montrose, Eitaro Aihara and Tongli Zhang**

Department of Pharmacology and Systems Physiology, University of Cincinnati College of Medicine, Cincinnati, OH, United States

Biological processes are dynamic. As a result, temporal analyses are necessary to fully understand the complex interactions that occurs within these systems. One example of a multifaceted biological process is restitution: the initial step in complex wound repair. Restitution is a dynamic process that depends on an elegant orchestration between damaged cells and their intact neighbors. Such orchestration enables the quick repair of the damaged area, which is essential to preserve epithelial integrity and prevent further injury. High quality dynamic data of the cellular and molecular events that make up the gastric restitution process has been documented. However, comprehensive dynamic models that connect all relevant molecular interactions to cellular behaviors are challenging to construct and experimentally validate. In order to efficiently provide feedback to ongoing experimental work, we have integrated dynamical modeling and machine learning to efficiently extract data-driven insights without incorporating detailed mechanisms. Dynamical models convert time course data into a set of static features, which are then subjected to machine learning analysis. The integrated analysis provides data-driven insights into how repair might be regulated in individual gastric organoids. We have provided a "*proof of concept*" of how such an analysis pipeline can be used to analyze any temporal dataset and provide timely data-driven insights.

Keywords: gastric epithelium, organoids, actin, restitution, computational model

OPEN ACCESS

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
Alessandro Giuliani,
Istituto Superiore di Sanità (ISS), Italy
Xiaojun Tian,
Arizona State University, United States

****Correspondence:***
Tongli Zhang
Tongli.Zhang@uc.edu;
zhangtl@ucmail.uc.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *27 April 2020*
Accepted: *24 July 2020*
Published: *12 August 2020*

Citation:
Ballweg R, Engevik KA, Montrose MH, Aihara E and Zhang T (2020) Extracting Insights From Temporal Data by Integrating Dynamic Modeling and Machine Learning. Front. Physiol. 11:1012. doi: 10.3389/fphys.2020.01012

INTRODUCTION

Temporal change is an essential feature of biological systems. An example of this is phenomena is found within the gastric epithelium. The epithelium closely regulates tissue integrity and rapidly responds to insult. When small injury or apoptosis occurs within the epithelial layer, dead or dying cells depart from the epithelial layer and neighboring cells immediately repair the damaged site by stimulating cell migration to close the gap (Xue et al., 2010; Xue et al., 2011; Aihara et al., 2013). This rapid repair response, known as restitution, does not involve proliferation and is sufficient to restore the epithelial barrier function (Ito et al., 1984; Lacy, 1995; Aihara et al., 2013; Aihara and Montrose, 2014). Restitution is a biologically important process to maintain epithelial integrity and prevent further expanding injuries. Numerous studies have identified effectors necessary for proper epithelial repair, including actin dynamics (Aihara et al., 2018). While fundamental repair mechanisms have been identified, the details behind these epithelial signaling cascades and sequences remain largely unclear.

Evaluating the mechanism behind restitution *in vivo* has proved difficult due to systemic limitations and limited tools to manipulate and monitor repair. While prior *in vitro* studies have identified potential pathways mediating repair, these findings largely rely on cancer derived cell line models which differ from the native tissue (Saeidnia et al., 2015). The introduction of the novel *in vitro* organoid model has allowed for growth of non-cancerous epithelial cells that contain cell types of the native tissues (Mahe et al., 2013; Schumacher et al., 2015a; Engevik et al., 2018). The organoid model has provided a three dimensional primary cell culture system which resembles the normal gastric epithelium, exhibiting similar gene expression patterns and functions as the native tissue (Schumacher et al., 2015a; Engevik et al., 2016; Aihara et al., 2018; Engevik et al., 2019). As a reductionist gastric restitution model, organoids can provide better insight to the innate epithelial cell response during damage in real time (Aihara et al., 2018; Engevik et al., 2019; Hanyu et al., 2019). *In vitro* organoid experimental methods produce high quality data sets that require extensive analysis to reveal underlying cellular events. As it is difficult to statistically compare two sets of dynamical data, current analysis is typically performed by comparing treatment groups at specific time points within the data, or specific features of the data such as a peak value that can be easily determined in each sample (Xue et al., 2010; Xue et al., 2011; Paek et al., 2016; Aihara et al., 2018; Engevik et al., 2019). The heterogeneity of organoids, which harbor multiple cell types, also adds to the complexity of analysis. Furthermore, detailed knowledge of the molecular control network is often unavailable for many biological systems, which generates a challenge if novel systems or pathways are studied. To address the above challenges, we have developed a novel analytical pipeline that converts single cell temporal microscopy data sets into data-driven, dynamic models that are then followed up with machine learning analysis. Here, we demonstrate the capability of this pipeline by using gastric epithelial repair as one example.

We have previously established gastric organoids, grown from epithelial stem cells of native mouse tissue, to investigate the innate epithelial restitution response (Schumacher et al., 2015a; Aihara et al., 2018; Engevik et al., 2018; Engevik et al., 2019). Using time course two-photon confocal microscopy of gastric organoids generated from human GFP-actin (HuGE) transgenic mice, we demonstrated that actin assembly occurs within the migrating cells neighboring the injured site, followed by recruitment of myosin II to provide cell contractility, which is regulated by RhoA/Rho associate kinase (ROCK) (Aihara et al., 2018). Actomyosin dynamics are particularly important in providing the force necessary to exfoliate the damaged cell away from the epithelial layer and to allow the neighboring cells to cover the denuded area (Stricker et al., 2010; Levayer and Lecuit, 2012; Aihara et al., 2018). While several actomyosin dynamic components and effectors have been identified as necessary for the role of actin in repair, the overall molecular mechanism remains unknown. Several studies have examined the molecular mechanisms of epithelial repair using computational models (Sherratt and Murray, 1990; Dale et al., 1994; Wearing and Sherratt, 2000; Tremel et al., 2009). Previously, we applied experimental data to a computational model and elucidated the role of force generated by actin dynamics during repair (Aihara et al., 2018).

To further connect molecular level events with outward cellular behaviors of this system, individual models were constructed to describe each step of the repair process (actin polymerization, dead cell movement and restitution) from experimental values detailed in Aihara et al. (2018). Aihara et al. (2018) applied a high power laser to gastric organoids, resulting in the damage and death of a single cell within an organoid (known as photodamage) (Section "Methods and Materials"). During subsequent high resolution imaging of the organoid, the role of actin was assessed over time by measuring actin polymerization and depolymerization (based upon actin GFP intensity) in neighboring cells. Additionally, measurements were made of the damaged area (the size of the damaged area based upon absence of GFP fluorescence in the damaged/dead cell), and dead cell distance (based upon the movement of the dead cell away from the damage site) (**Figure 1B**; Aihara et al., 2018). Using these experiment-derived parameters (Aihara et al., 2018), which summarize the dynamical characteristics of individual gastric organoids, we assessed how different treatments affect system dynamics. Random forest analysis of these data-derived parameters made clear the dominant features that link the molecular level events to the cellular level behaviors. Our analysis provides deep insights on how the molecular and cellular events control damage repair, which will move forward the development of targeted therapeutics for the native tissue repair.

Since our pipeline can be applied to analyze individual time course data sets to extract insights across various systems to enhance our biological understanding, we expect that our analysis pipeline can be a useful tool to provide efficient analysis and promote effective interdisciplinary collaboration.

RESULTS

Extracting Cell-Specific Features From Dynamical Data

Restitution is an intrinsic function in gastric epithelial cells and generally involves the following events: (1) orderly reorganization of cell-cell and cell-substratum contact to allow viable cells to migrate from their site of origin, (2) cytoskeletal reorganization, including the formation of lamellipodia to promote cell movement, and (3) signal production to promote dead cell exfoliation, viable cell migration and prevent cell death within the viable cells (Aihara et al., 2017; Aihara et al., 2018). All steps are essential to restore epithelial barrier function (**Figures 1A,B**). Time course microscopy of high power laser induced single cell damage and subsequent repair suggests that the timing of these events is integral to the overall temporal trajectory for each measurement: actin GFP-intensity, dead cell distance, and damaged area (**Figures 1C–E**). In GFP-actin gastric organoid experiments, the time at which actin reaches minimal fluorescence ($\boldsymbol{T_{min}}$), indicates the elimination of actin within the damaged cell and the initial time where F-actin begins to accumulate within the neighboring cells (**Figure 1C**). This initial decay is controlled by the parameter ($\boldsymbol{Act_{depoly2}}$). The time at

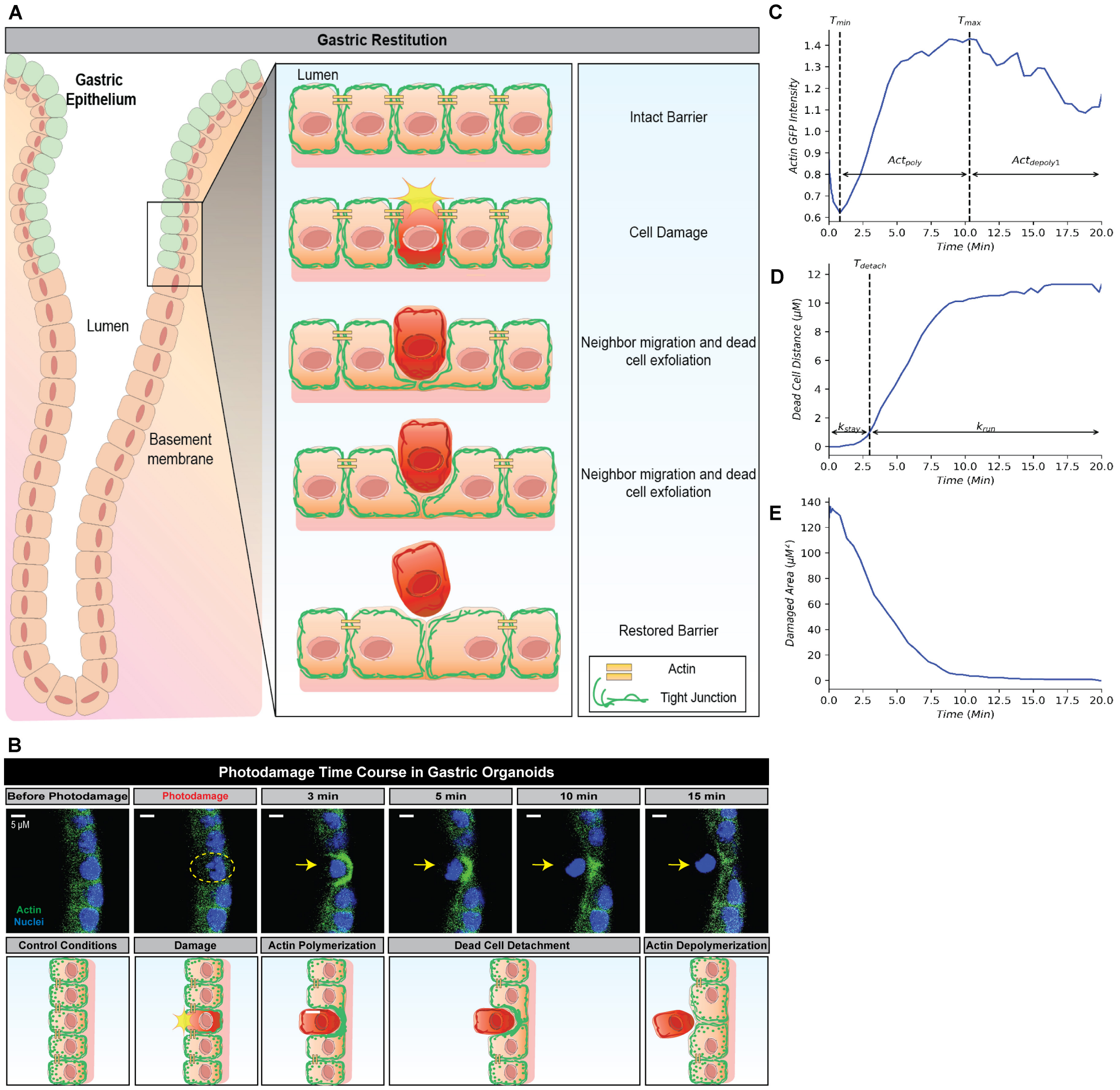


FIGURE 1 | Simplified framework to recapture gastric repair dynamics. **(A)** Illustration of gastric epithelial repair response to single cell damage *in vivo*. Following cell death or damage, intact cells neighboring the damaged cell undergo actin polymerization which promotes migration toward the area of damage in order to fill the denuded area and assist in exfoliating the damaged cell away from the epithelium. Through this actin polymerization and subsequent migration, the neighboring cells ensure there are no denuded areas and restore continuity to the epithelial barrier. *In vivo*, these steps are important to promote epithelial homeostasis and prevent further damage to the epithelium and underlying mucosal layers. **(B)** Representative image (above) and illustration (below) of epithelial repair response to single cell damage *in vitro* in gastric organoids. Following single cell damage by high power laser (known as photodamage), the damaged cell (above; indicated by yellow dashed circle and arrow) loses GFP-fluorescence while intact neighboring cells show increased GFP-actin in green (indicating actin polymerization) in the lateral region closest to the damage site. The neighboring cells migrate and fill in the damage area, followed by the exfoliation of the damaged cell toward the lumen (dead cell detachment). This results in actin depolymerization and ultimately a restored epithelial barrier function. **(C)** Temporal trajectory of actin following damage (t = 0 min), actin first begins to degrade to a minimum where it then accumulates to a maximum level (actin polymerization) before degrading to a steady state (actin depolymerization). T_{min} and T_{max} indicate the time at which actin reaches the minimum and maximum values, respectively. **(D)** Temporal trajectory of the dead cell following damage (t = 0). The cell begins to move away at a constant rate before detaching from its neighboring cells; following detachment, the dead cell rapidly moves to a maximum distance from the damage site. T_{detach} indicates the time of detachment. **(E)** Temporal trajectory of the damaged area following damage (t = 0). Following single cell damage, the maximum damage size is reached at t = 0 and decreases in size overtime indicating repair.

which actin reaches peak fluorescence (T_{max}) determines when actin begins to contract, assuming that F-actin dissociates into G-actin at this point while the sustained increase of fluorescence likely occurs due to cellular contraction (**Figure 1C**). The increase of actin to its maximum is controlled by the parameter Act_{poly}. The decrease of actin to its steady state (Act_{SS}) is controlled by the parameter $Act_{depoly2}$. During the repair process, dead cell exfoliation occurs following detachment from the neighboring cells (T_{detach}, **Figure 1D**), where upon the damaged cell moves rapidly to a maximal distance (dead cell distance, DCD_{max}) from the monolayer that contains all the viable cells. The rate controlling this movement is defined as k_{run}. Additionally, the damaged area is repaired in a pattern that follows an exponential decay curve controlled by the parameter T_{repair} (**Figure 1E**).

Due to the heterogeneous nature of the cells within gastric organoids, which reflect the gastric epithelium, each experiment results in a different time dependent trajectory for the three characteristics described above. To properly align these cellular and molecular events into a time dependent chain of events, we chose to use a hybrid model (section "Methods and Materials"). The hybrid model best describes the dynamics of each event as the repair process involves continuous time dependent changes and discrete timing events (Bartocci and Lio, 2016). Hence, these dynamical models decompose the time dependent microscopy data for each individual experiment into a collection of cell-specific features which describe the dynamics observed as a single train of events for the cell under study.

Assembling Physiological Repair Data Within a Single Dynamical Framework

Following the development of the data-driven models, we tested whether this modeling framework could recapture the temporal dynamics of gastric repair within control experiments. Both the model and data indicated that actin dynamics undergo a three-stage response to cellular damage: actin polymerization and depolymerization (actin-GFP intensity), decrease of damage area (damaged area), and exfoliation of the dead cell from the epithelial monolayer (dead cell distance). In the initial stage, actin within the damaged cell diminishes (indicated by loss of GFP fluorescence, **Figure 1B**) at an unknown rate before it reaches a minimal level. This rate is unknown as this rapid event cannot be captured following damage due to the imaging time course. Once the minimal level is reached, actin polymerization increases in the neighboring cells followed by these neighboring cells migrating to cover the damage area. During this set of events, actin polymerization arrives at a maximal (peak) level. Following its peak, actin decreases (actin depolymerization) until it reaches steady state. Using our model, we simulated the actin trajectory at a single cell level for more than 20 experiments (blue dashed lines **Figure 2A** and **Table 1**). The simulated dynamics agreed reasonably well with the measured actin levels (blue dots **Figure 2A**) and recaptured all three stages of actin's response to damage (actin polymerization, peak, and actin depolymerization). Simulating the average actin dynamics (black line **Figure 2A**) showed that the average model simulation fell within the region bounded by the experimentally observed actin levels, indicating that the overall model fit was reasonable.

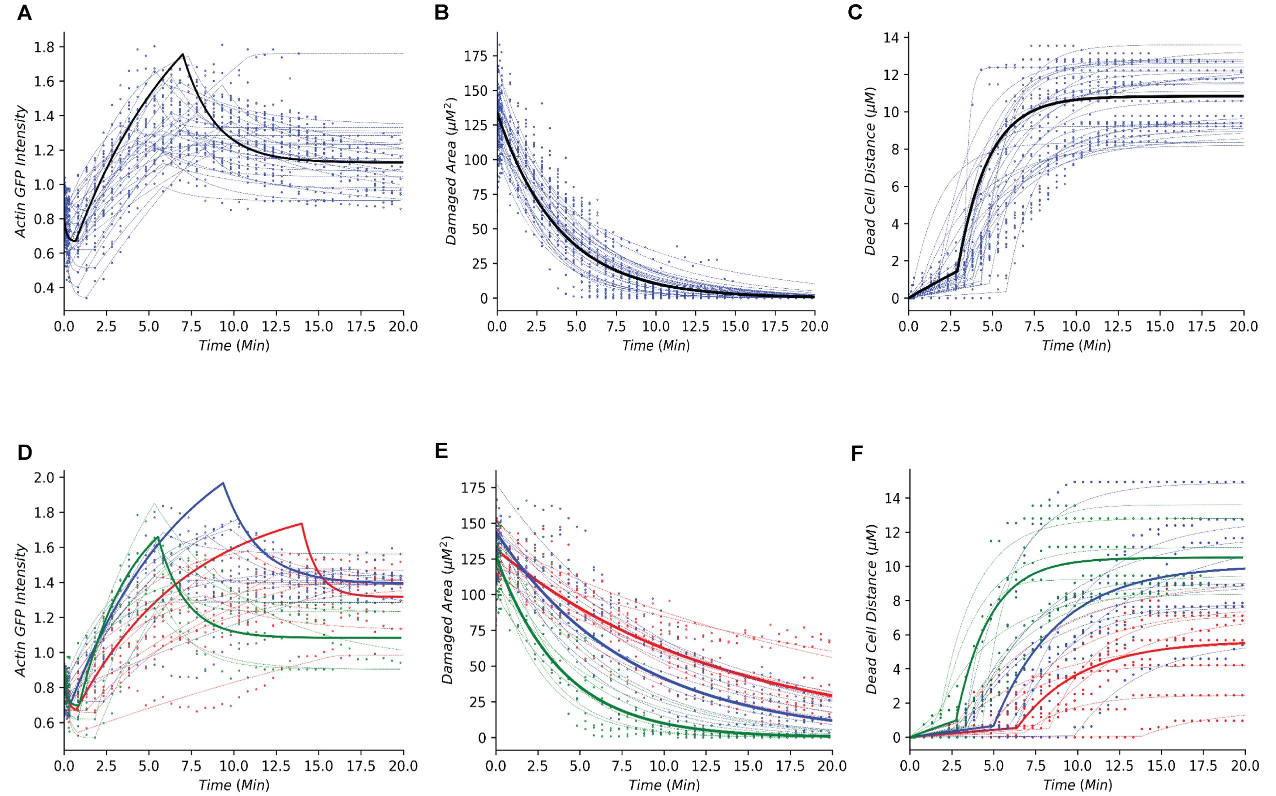


FIGURE 2 | Simple dynamical models recapture the gastric repair process in control and perturbed scenarios. **(A–C)** Temporal profiles of actin, damaged area, and dead cell movement, respectively. Blue dots represent experimental data from control (untreated) organoids that have been previously reported in Aihara et al. (2018). Blue dashed lines represent model simulations for individual organoid outcomes, and black lines represent the simulation of an average control cell. **(D–F)** Temporal profiles of actin, damaged area, and dead cell movement in control (green), or treatment with Blebbistatin at high dosage (red) or low dosage (blue). Dots represent data collected from the individual experiments in Aihara et al. (2018). Dashed lines represent model simulations and solid lines show the simulation of an average cell.

TABLE 1 | Average parameter values and respective units from control organoid experiments shown in **Figures 2A–C**.

Parameter	Value (Mean ± SD)	Units
$\text{ACt}_{depolyl}$	6.024 ± 3.557	min^{-1}
$\text{Act}_{depoly2}$	0.616 ± 0.39	min^{-1}
ACt_{poly}	0.191 ± 0.171	min^{-1}
ACt_{max}	2.484 ± 0.944	RLU
ACt_{min}	0.696 ± 0.115	RLU
Act_{ss}	1.083 ± 0.168	RLU
DCD_{max}	10.507 ± 2.03	μm
K_{repair}	0.255 ± 0.08	μm^2 min^{-1}
K_{run}	0.462 ± 0.287	μm min^{-1}
K_{stay}	0.035 ± 0.023	μm min^{-1}
T_{detach}	2.749 ± 1.134	Min
T_{max}	5.534 ± 1.185	Min
T_{min}	0.886 ± 0.726	Min

Experimental data indicated that repair of the damaged area, as indicated by the decrease in damage size over time, is a single-stage process that appears to follow an exponential decay curve (blue dots, **Figure 2B** and **Table 1**). With our model, we simulated the repair of the damaged area at a single cell level for more than 20 experiments (blue dashed lines **Figure 2B**). The simulated repair of the damaged area agreed reasonably well with the experimentally observed repair (blue dots **Figure 2B**). Furthermore, simulation of the average repair of the damaged area (black curve **Figure 2B**) we observed that the overall model fit was able to recapture the experimental observations.

In addition to actin dynamics and repair of the damaged area, the experimental data also indicated that dead cell exfoliation (dead cell distance) is a two-stage process whereupon the dead cell is slowly moved away from the monolayer at a constant rate and then at a specific time ($\mathbf{T}_{detach}$) the cell is fully detached and the dead cell accelerates until it reaches steady state distance from the damaged area (blue dots, **Figure 2C** and **Table 1**). Using our model, we simulated the distance traveled in each experiment (blue dashed lines, **Figure 2C**). The model was able to recapture the initial constant velocity, acceleration and steady state of the experimental observation. The average simulated dead cell was well within the region bounded by the experimental results indicating that the derived parameters were reasonable (black line, **Figure 2C**).

The ability of our modeling framework to recapture these key features of normal gastric repair suggests that the model assumptions and the derived parameter values are reasonable. With the affirmation that our model assumptions align with the cellular events under normal conditions, we then examined whether the same approach could be used to study perturbation experiments.

Organization of the Perturbed Dynamics Within the Dynamical Framework

For perturbation studies of the model, we chose a common myosin inhibitor (Blebbistatin) which shows a clear dose dependent response on the temporal dynamics of gastric repair (**Figures 2D–F**; Aihara et al., 2018). Using the same data-driven approach (Section "Methods and Materials"), we fit single cell models for the control experimental group (green dots **Figures 2D–F**), along with high dosage and low dosage Blebbistatin experimental groups (red and blue dots, respectively, **Figures 2D–F**). Our models faithfully recaptured the actin dynamics, repair of damage, and dead cell distance of all three treatment groups of control, high dosage Blebbistatin and low dosage Blebbistatin (dashed lines, **Figures 2D–F**). As in the control case, average cells (thick lines, **Figures 2D–F**) were well within the regions bound by their respective data sets, indicating that the derived parameters were within reason. The model was also able to recapture the dose dependent response of the Blebbistatin treatments. For instance, a low dose of Blebbistatin causes a decrease in repair compared to control experiments, which is further decreased by a higher dose of Blebbistatin. This phenomenon was also observed in the maximal distance reached by the dead cell. In concordance with Blebbistatin, the computational framework was able to recapture the remaining experimental data published by Aihara et al. (2018). As the model was able to recapture such a wide array of experimental scenarios, this method could be considered a solid framework to study the gastric repair process *in silico* and additional analysis of the derived parameters can offer mechanistic insights into the key events that characterize the restitution process.

Chord Plotting Reveals the Dominant Effect of the Perturbations

Following simulation for each treatment group, we aggregated both the model derived and data-driven parameters to look for any additional insight into the potential effect of various inhibitors previously tested (Aihara et al., 2018) on the events that control the gastric repair process. For each treatment group, the percent change between the model parameters of the treatment and control experiments was calculated. The parameters exhibiting the highest positive change and the parameters values with the highest negative change were then plotted on a chord plot (**Figure 3**), which connects each treatment with the mechanistic properties they most strongly regulate (increases indicated by green arrows; decreases indicated by red arrows). By plotting interactions in this form, we provide a concise picture into the effect of each inhibitor during restitution. For instance, NSC23766 (a selective Rac1 inhibitor) causes a large increase in k_{repair} (green arrow, **Figure 3A**), indicating that the treatment causes an increase in the rate of damage repair when compared to controls. In contrast, ML-141 (a potent, selective inhibitor of Rho family GTPase cdc42) decreases the rate of gastric repair, while also increasing T_{max}, (green arrow, **Figure 3A**), indicating that the rate of gastric repair may be associated with an increase in T_{max} compared to controls. ML-141 has been shown to decrease actin polymerization in airway epithelial cells (Ferru-Clement et al., 2015). Furthermore, we observe that both dosages of Blebbistatin affect the same mechanistic control parameters (k_{stay} and T_{detach}), consistent with the dose dependent response of this treatment shown in **Figures 2D–F**.

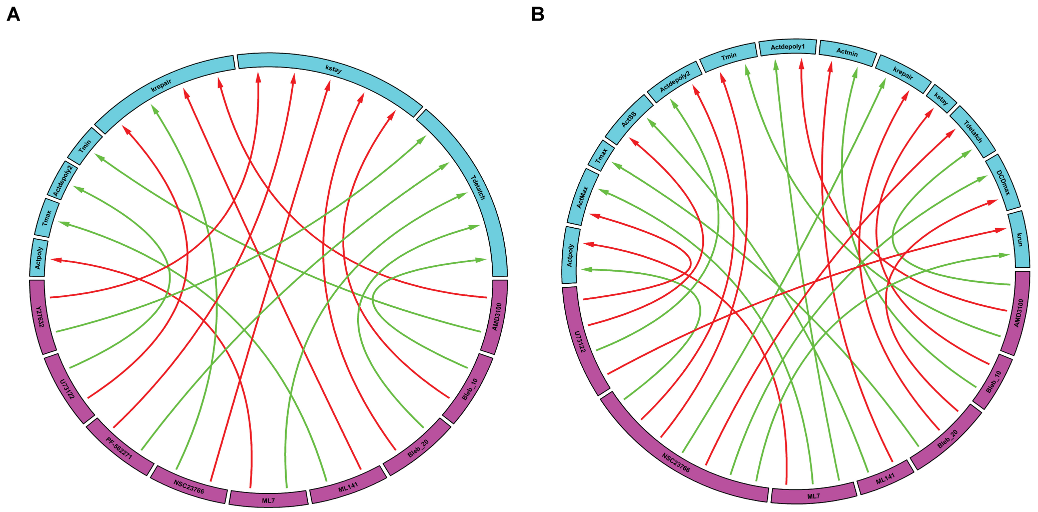


FIGURE 3 | Data-driven analysis provides insights into the control of gastric repair. Arrows originate from the drugs (pink boxes) and point to a feature of the model (blue boxes). **(A)** Chord plot of the maximal interactions of each drug treatment to model parameters. Each drug can inhibit or activate a model feature. A green arrow represents the parameter that is most activated by a drug. A red arrow indicates the feature that is most inhibited by a drug. **(B)** Chord plot of the maximal enhancement or inhibition of each model feature by any drug treatment. Each model feature (pink boxes) can be inhibited or activated by a drug treatment (blue boxes). A red arrow indicates that the treatment has the highest inhibition of the model feature when compared to all other treatments. A green arrow indicates that the treatment has the highest activation of a model feature when compared to all other treatments. It is important to note that there are differences in the parameters and treatments displayed when comparing figures, **(A,B)**. In **(A),** not all parameters are shown as some might not be greatly affected by the treatments being examined; whereas in **(B),** not all treatments are shown as they might not have large effects on the model parameters when compared to the rest of the treatments.

We also plotted the overall maximum enhancement or inhibition of each model parameter (**Figure 3B**) compared to treatment. Each parameter contains two inputs: the maximum enhancement effect and the maximum inhibitory effect. Note that T_{max} and k_{stay} only have a single input as all treatment groups were shown to increase T_{max} or decrease (k_{stay}) control parameters. From this plot, we can deduce which drugs have the greatest overall effect (compared to all other drugs) on any of the model's features. For instance, AMD3100 (a CXCR4 inhibitor) showed the largest overall enhancement effect on T_{detach} (the parameter controlling the time at which the dead cell loses connection with its neighbors). Additionally, AMD3100 had the largest inhibitory effect compared to all other drugs on $Act_{depoly1}$ (parameter for actin depolymerization). Based on this data and our observations of the effects of drugs on individual parameters, we are able to observe which treatments have the most profound effects on the overall repair mechanism. Interestingly, NSC23766 caused the most changes to the model features, with a total of 7 connections, indicating that NSC23766 might have the most profound effect on the overall system compared to all other treatments in the plot. On the basis of such results, we can then identify the role of Rac1 (inhibited by NSC23766) as a promising target during repair for future studies.

Random Forest Reveals the Control of Damage Repair and Dead Cell Exfoliation by Actin

The previous analysis demonstrated the effect of individual treatment on actin accumulation, repair of the damaged area, and dead cell exfoliation, but does not address the potential connections between these processes. To investigate how the events can be connected, we then analyzed whether the molecular level events (actin polymerization and depolymerization) might be associated with the cellular level behaviors (damage repair, cell exfoliation). While prior analysis separated experiments into groups based upon treatment scenarios, we carried out a pooled study whereupon all experiments (treated and controls) are combined into a single population which exhibited heterogeneous actin dynamics. To assess the correlation amongst the model parameters we performed a principal component analysis (PCA). A biplot of the analysis (**Figure 4A**) points to a multitude of correlations between the parameters controlling actin dynamics and parameters controlling cellular behaviors. In particular, there is a strong positive correlation between $\mathbf{T}_{max}$ and $\mathbf{T}_{detach}$. Additionally, there were strong negative correlations between **Tmax** and parameters controlling dead cell exfoliation ($\mathbf{k}_{run}$, $\mathbf{DCD}_{max}$) and the parameter controlling the repair of the damaged site ($\mathbf{k}_{repair}$).

The relationship between the parameters which describe actin dynamics and the rate at which the damaged site is repaired were further assessed using a tree-based regression method (Random Forest regression). An ensemble of regression trees was built with the data-driven parameters and model derived rates for actin, which we then used to predict the corresponding restitution rates from both a training and validation cohort (section "Methods and Materials"). The random forest performed well in predicting restitution rates for both training and validation cohorts (R^2 Training = 0.93 and R^2 Validation = 0.34) indicating a robust statistical model which could be used for future study. Due to

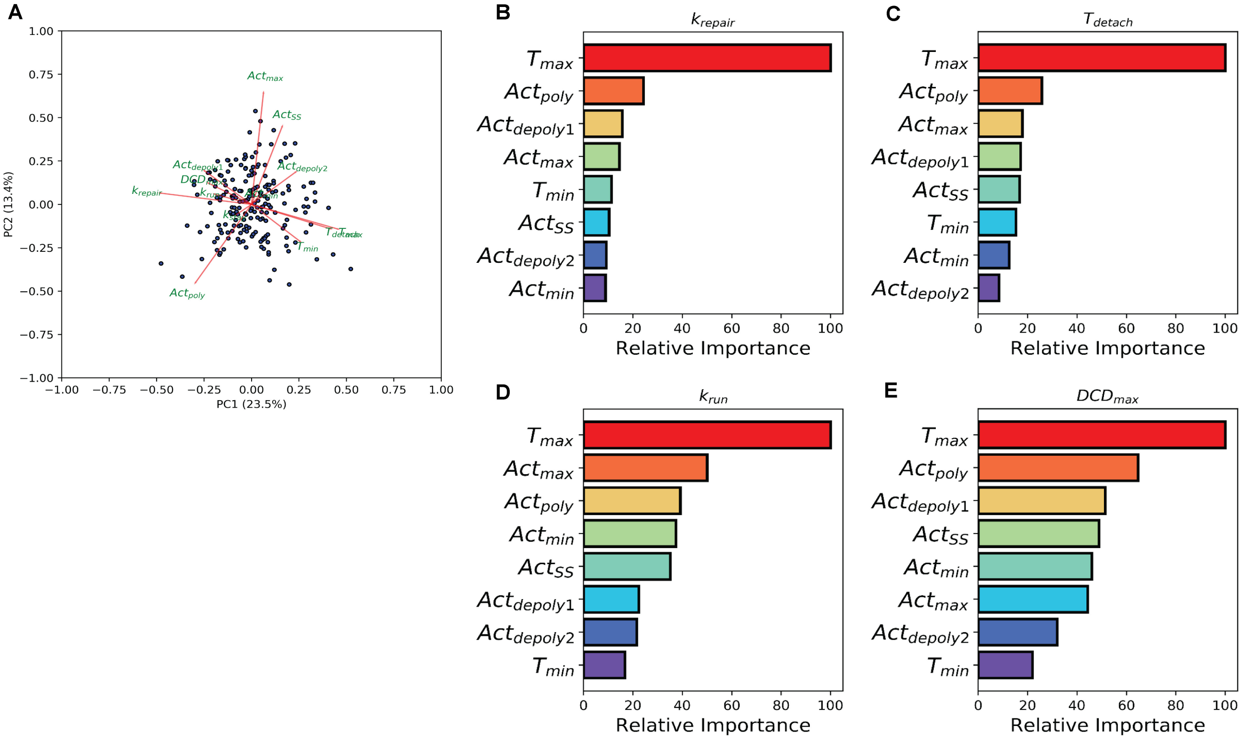


FIGURE 4 | Machine learning links timing events between molecular and cellular level events. **(A)** PCA analysis biplot of the model parameters. Dots represent samples plotted based on the first two principal components. Red arrows represent the loadings from the PCA. **(B–E)** Bar plots depicting Random Tree variable importance analysis for the indicated parameters: **(B)** repair rate (k_{repair}), **(C)** detachment time (T_{detach}), **(D)** rate at which the dead cell is pushed out by its neighbors (k_{run}), and **(E)** maximum dead cell distance (DCD_{max}). Axis is scaled 0 – 100, where 100 represents parameters that are highly significant in the random forest model and 0 represents insignificant parameters.

the strength of the Random Forest Regression, we performed a variable importance analysis to identify which actin parameters made significant contributions to the statistical model. This analysis revealed that T_{max} (the time at which actin reaches its peak) plays a significant role in determining the rate of repair (**Figure 4B**), suggesting a potential mechanistic link between the two processes. The random forest analysis was then repeated for each of the parameters describing dead cell extrusion (k_{stay}, T_{detach}, k_{run}, and DCD_{max}). Interestingly, T_{max} was revealed to be important for T_{detach}, k_{run} and DCD_{max} (**Figures 4C–E**); indicating that the time of peak actin accumulation is a strong predictor for the dynamics of the cell exfoliation process. Hence without explicitly incorporating any knowledge, our data-driven analysis reveals a critical role for actin polymerization timing in determining the cellular level behaviors involved in the gastric repair process.

DISCUSSION

In this work, we have developed a computational pipeline that can efficiently examine key factors involved in *in vitro* restitution using temporal data collected from time lapse confocal microscopy of gastric organoids (Aihara et al., 2018). Our approach reveals that the timing of actin cytoskeleton rearrangement greatly influences the repair of the damaged area, as well as the exfoliation of the damaged cell. We predict that these functions are critical for the healing process and have implications in situations, such as *Helicobacter pylori* infection, where wound healing is delayed (Aihara et al., 2014; Hanyu et al., 2019). *In vitro* organoid studies in epithelial repair (Aihara et al., 2018; Engevik et al., 2019; Engevik et al., 2020) have been demonstrated to be similar to native tissue as they demonstrate (1) importance of actin dynamics in cells neighboring damage sites (Aihara et al., 2018), (2) increased calcium mobilization, which is dependent upon trefoil factor 2 signaling during repair (Xue et al., 2010), and (3) role of sodium hydrogen exchanger 2 downstream of the trefoil factor 2 pathway in repair (Xue et al., 2011). These *in vitro* studies identified features intrinsic to the epithelium and is reflective in native tissue. Utilization of data sets from organoid experiments, which exhibit similar responses to infection and damage as native tissue (Schumacher et al., 2015a,b; Aihara et al., 2018), provides a reductionist platform to study repair at the cellular level with the potential to better translate to *in vivo* events.

Current methods of analysis for epithelial repair, such as exponential analysis of the damaged site or comparison between specific time points, offer statistical differences between experimental groups (Aihara et al., 2018) and indicate important aspects of repair. However, these techniques have not taken full advantage of all the temporal data or revealed the underlying heterogeneity between individual cells. On the other hand, though mechanistic modeling is suitable for dynamical data sets and numerous mathematical models of varying complexity have indeed provided valuable insights on epithelial repair and/or wound repair (Dale et al., 1994; Wearing and Sherratt, 2000), the development of these models often requires significant amount of time and resource.

To facilitate theory-experiment collaboration with timely feedbacks, we have proposed to integrate simple dynamical models with following up machine learning analysis. This approach allowed us to utilize the entire data set while incorporating minimal assumptions (time dependent behaviors for each process and that all processes are independent) about each process. By using this data-driven approach, the model parameters are derived directly for each individual cell. In doing so, each cell can be encoded as a group of dynamic features that have been extracted directly from data and are associated with each other during analysis. This method decomposes the time course data into static features which are suitable for downstream analysis. Machine learning algorithms such as random forest regression can then be used to compare differences between individual cells and reveal what features are vital for the gastric restitution process. This combination provides the foundation to look further into the cellular and molecular events essential for proper repair.

Biological systems are very complex and proper understanding of them is beyond the capability of any single method, and we believe the new area of systems biology calls for the synergistic integration of all available tools. In our example we have shown that a proper combination of real time microscope, cutting edge organoid culture, dynamical modeling, and machine learning can produce efficient insights on a new biological system. We believe that such combination can be applicable to any biological systems and will facilitate deep understanding of them.

MATERIALS AND METHODS

Experimental Data Acquisition

The experimental data used for this study has previously been published (Aihara et al., 2018). Gastric organoids were generated from HuGE (Human GFP-Actin Expressing) transgenic mice as previously described (Schumacher et al., 2015a; Aihara et al., 2018). Using a two-photon confocal microscope (Zeiss LSM 510 NLO), images of gastric organoid nuclei (Hoechst 33342; Ti-Sa excitation 730 nm, emission 435–485 nm) and actin (GFP; excitation 488 nm, emission 500–550 nm) were collected (Aihara et al., 2018). After collecting a set of control images, a small region (∼5 μm^2) of a cell within the gastric organoid was repetitively scanned for 2–3 s at high laser power resulting in single cell damage (photodamage) (Aihara et al., 2018). For inhibitory experiments, gastric organoids were incubated with drugs at least 1 h prior to experiments: AMD3100 (1 μM, Sigma), U73122 (10 μM, Enzo life sciences), ML7 (10 μM, Calbiochem), Blebbistatin (20 μM or 10 μM, Sigma), NSC23766 (50 μM, Cayman), PF-562271 (1 μM, gift from Dr. James E. Casanova, Univ Virginia), ML141 (20 μM, Calbiochem), and Y27632 (20 μM, Enzo Life Sciences). The parameters measured over time from the collected gastric organoid images include: actin polymerization based upon GFP intensity (Aihara et al., 2018; Hanyu et al., 2019), size of damage area (Aihara et al., 2018; Engevik et al., 2019; Hanyu et al., 2019; Engevik et al., 2020), and dead cell distance based upon movement of damage cell nuclei away from the site of damage (Aihara et al., 2018; Engevik et al., 2019; Hanyu et al., 2019). The damage-repair cycle was measured once per gastric organoid, and outcomes of at least four different gastric organoids were compiled for each experiment.

Model Development

To recapture each of the experimentally observed features we use a collection of piecewise linear ordinary differential equations (ODE). Each feature is modeled as an independent process and is controlled by a single ODE.

Actin Model

$$\frac{dAct}{dt} = Act_{poly} * (Act_{max} - Act) - Act_{depoly2} * (Act - Act_{SS}) - Act_{depoly1} * (Act - Act_{min})$$

Where Act_{poly} represents the rate of actin polymerization, which is only on between Act_{max} the maximum value of actin, $Act_{depoly2}$ the rate at which actin degrades following the actin peak, Act_{SS} the steady state that actin degrades to. Act_{min} is the minimal actin level for each cell and $Act_{depoly1}$ is the rate that actin degrades to Act_{min}. Rate constants are turned on or off dependent on certain timing events with: $Act_{depoly1}$ on between $t = 0$ and $t = T_{min}$; Act_{poly} on between $t = T_{min}$ and $t = T_{max}$ and $Act_{depoly1}$ on between $t = T_{max}$ and the end of the simulation.

Damaged Area Model

$$\frac{dDA}{dt} = - k_{repair} * DA$$

Where k_{repair} represents the rate of repair of the damaged area.

Dead Cell Distance

$$\frac{dDCD}{dt} = k_{stay} + k_{run} * (DCD_{max} - DCD)$$

Where k_{stay} represents the background force acting on the dead cell, k_{run} is the rate at which the dead cell is pushed out by neighboring cells and DCD_{max} is the maximum distance traveled by the dead cell. k_{run} is turned on when the dead cell loses connection with its neighboring cells (T_{detach}).

Model Parameterization

Parameters controlling the timing events (T_{max}, T_{min}, and T_{detach}) along with Act_{min} and Act_{SS} were extracted directly from the data for individual organoids in all treatment groups. To extract rate constants and Act_{max} (assumed to be theoretical), the ODEs were first separated into different events and integrated with respect to time. An evolutionary algorithm was then used to minimize the sum of squares between model simulation for each experiment from Aihara et al. (2018). This process was repeated for each experiment, generating a cohort of parameter sets, each representing an individual organoid.

Principal Component Analysis

To understand the potential relation among individual experiments and correlation among derived model parameters in an unsupervised approach, we performed a Principal Component Analysis (PCA). All of the derived parameters were considered and scaled prior to analysis. The PCA was performed using the PCA function within the scikit-learn package.

Random Forest Analysis

To understand how the parameters that control actin influence the cellular behaviors (Dead Cell Movement, Repair) we used Random Forest regression, an ensemble tree based regression method. This method was chosen as it does not require knowledge of how the predictors should be combined. The actin parameters for all groups (treatment and controls) were used as predictors while the parameters controlling cell behaviors (k_{repair}, T_{detach}, k_{stay}, k_{run}, and DCD_{max}) were used as response variables. The data was then randomly split into Training and Validation sets and the Random Forest model was built using the default settings contained within the scikit-learn package. A variable importance analysis was performed to assess the importance of each predictor (parameters controlling actin dynamics) on the response variables (parameters controlling damage repair or cell exfoliation). This was done using the default calculation given within the scikit-learn package, which assesses the Gini importance of each predictor. The importance values were then scaled to the maximal importance and compared.

Packages Used

All ordinary differential equations were implemented in Python and solved using ODEint from the scipy package[1]. Chord plots were generated using Circlize (Gu et al., 2014) implemented in R^2. Random forest regression was done using the standard package implemented in scikit-learn[3].

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available from the authors upon request.

AUTHOR CONTRIBUTIONS

RB, KE, MM, EA, and TZ conceived and designed the research. RB, KE, EA, and TZ performed the experiments and analyzed the data. All authors edited and revised the manuscript. MM, EA, and TZ obtained research funding for experiments and computational modeling.

FUNDING

This work was supported by the National Institutes of Health (NIH) grant R01DK102551 (MM, EA), F31DK115126 (KE), R21CA227379 (TZ), and Ryuji Ueno Award co-sponsored by the S&R Foundation and American Physiological Society (EA).

---

[1] https://www.scipy.org/

[2] https://www.r-project.org/

[3] https://scikit-learn.org/stable/

REFERENCES

Aihara, E., Closson, C., Matthis, A. L., Schumacher, M. A., Engevik, A. C., Zavros, Y., et al. (2014). Motility and chemotaxis mediate the preferential colonization of gastric injury sites by *Helicobacter pylori*. *PLoS Pathogens* 10:e1004275. doi: 10.1371/journal.ppat.1004275

Aihara, E., Engevik, K. A., and Montrose, M. H. (2017). Trefoil factor peptides and gastrointestinal function. *Annu. Rev. Physiol.* 79, 357–380. doi: 10.1146/annurev-physiol-021115-105447

Aihara, E., Hentz, C. L., Korman, A. M., Perry, N. P., Prasad, V., Shull, G. E., et al. (2013). In vivo epithelial wound repair requires mobilization of endogenous intracellular and extracellular calcium. *J. Biol. Chem.* 288, 33585–33597. doi: 10.1074/jbc.m113.488098

Aihara, E., Medina-Candelaria, N. M., Hanyu, H., Matthis, A. L., Engevik, K. A., Gurniak, C. B., et al. (2018). Cell injury triggers actin polymerization to initiate epithelial restitution. *J. Cell Sci.* 131:jcs216317. doi: 10.1242/jcs.216317

Aihara, E., and Montrose, M. H. (2014). Importance of Ca(2+) in gastric epithelial restitution-new views revealed by real-time in vivo measurements. *Curr. Opin. Pharmacol.* 19, 76–83. doi: 10.1016/j.coph.2014.07.012

Bartocci, E., and Lio, P. (2016). Computational modeling, formal analysis, and tools for systems biology. *PLoS Comput. Biol.* 12:e1004591. doi: 10.1371/journal.pcbi.1004591 doi: 10.1371/journal.pcbi.1004591

Dale, P. D., Maini, P. K., and Sherratt, J. A. (1994). Mathematical modeling of corneal epithelial wound healing. *Math. Biosci.* 124, 127–147. doi: 10.1016/0025-5564(94)90040-x

Engevik, A. C., Feng, R., Choi, E., White, S., Bertaux-Skeirik, N., Li, J., et al. (2016). The development of spasmolytic polypeptide/tff2-expressing metaplasia (spem) during gastric repair is absent in the aged stomach. *Cell Mol. Gastroenterol. Hepatol.* 2, 605–624. doi: 10.1016/j.jcmgh.2016.05.004

Engevik, K. A., Hanyu, H., Matthis, A. L., Zhang, T., Frey, M. R., Oshima, Y., et al. (2019). Trefoil factor 2 activation of CXCR4 requires calcium mobilization to drive epithelial repair in gastric organoids. *J. Physiol.* 597, 2673–2690. doi: 10.1113/jp277259

Engevik, K. A., Karns, R. A., Oshima, Y., and Montrose, M. H. (2020). Multiple calcium sources are required for intracellular calcium mobilization during gastric organoid epithelial repair. *Physiol. Rep.* 8:e14384.

Engevik, K. A., Matthis, A. L., Montrose, M. H., and Aihara, E. (2018). Organoids as a model to study infectious disease. *Methods Mol. Biol.* 1734, 71–81. doi: 10.1007/978-1-4939-7604-1_8

Ferru-Clement, R., Fresquet, F., Norez, C., Metaye, T., Becq, F., Kitzis, A., et al. (2015). Involvement of the Cdc42 pathway in CFTR post-translational turnover and in its plasma membrane stability in airway epithelial cells. *PLoS One* 10:e0118943. doi: 10.1371/journal.pone.0118943

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393

Hanyu, H., Engevik, K. A., Matthis, A. L., Ottemann, K. M., Montrose, M. H., and Aihara, E. (2019). *Helicobacter pylori* uses the TlpB receptor to sense sites of gastric injury. *Infect. Immun.* 87, e00202–e00219.

Ito, S., Lacy, E., Rutten, M., Critchlow, J., and Silen, W. (1984). Rapid repair of injured gastric mucosa. *Scand. J. Gastroenterol. Suppl.* 101, 87–95.

Lacy, E. R. (1995). Rapid epithelial restitution in the stomach: an updated perspective. *Scand. J. Gastroenterol.* 30, 6–8. doi: 10.3109/00365529509090260

Levayer, R., and Lecuit, T. (2012). Biomechanical regulation of contractility: spatial control and dynamics. *Trends Cell Biol.* 22, 61–81. doi: 10.1016/j.tcb.2011.10.001

Mahe, M. M., Aihara, E., Schumacher, M. A., Zavros, Y., Montrose, M. H., Helmrath, M. A., et al. (2013). Establishment of gastrointestinal epithelial organoids. *Curr. Protocols Mouse Biol.* 3, 217–240. doi: 10.1002/9780470942390.mo130179

Paek, A. L., Liu, J. C., Loewer, A., Forrester, W. C., and Lahav, G. (2016). Cell-to-cell variation in p53 dynamics leads to fractional killing. *Cell* 165, 631–642. doi: 10.1016/j.cell.2016.03.025

Saeidnia, S., Manayi, A., and Abdollahi, M. (2015). From in vitro experiments to in vivo and clinical studies; pros and cons. *Curr. Drug. Discov. Technol.* 12, 218–224. doi: 10.2174/1570163813666160114093140

Schumacher, M. A., Aihara, E., Feng, R., Engevik, A., Shroyer, N. F., Ottemann, K. M., et al. (2015a). The use of murine-derived fundic organoids in studies of gastric physiology. *J. Physiol.* 593, 1809–1827. doi: 10.1113/jphysiol.2014.283028

Schumacher, M. A., Feng, R., Aihara, E., Engevik, A. C., Montrose, M. H., Ottemann, K. M., et al. (2015b). *Helicobacter pylori*-induced sonic hedgehog expression is regulated by NFkappaB pathway activation: the use of a novel in vitro model to study epithelial response to infection. *Helicobacter* 20, 19–28. doi: 10.1111/hel.12152

Sherratt, J. A., and Murray, J. D. (1990). Models of epidermal wound healing. *Proc. R. Soc. Lond. B.* 241, 29–36.

Stricker, J., Falzone, T., and Gardel, M. L. (2010). Mechanics of the F-actin cytoskeleton. *J. Biomech.* 43, 9–14.

Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B., Stevens, G., Landman, K., et al. (2009). Cell migration and proliferation during monolayer formation and wound healing. *Chem. Eng. Sci.* 64, 247–253. doi: 10.1016/j.ces.2008.10.008

Wearing, H. J., and Sherratt, J. A. (2000). Keratinocyte growth factor signalling: a mathematical model of dermal–epidermal interaction in epidermal wound healing. *Math. Biosci.* 165, 41–62. doi: 10.1016/s0025-5564(00)00008-0

Xue, L., Aihara, E., Podolsky, D. K., Wang, T. C., and Montrose, M. H. (2010). In vivo action of trefoil factor 2 (TFF2) to speed gastric repair is independent of cyclooxygenase. *Gut* 59, 1184–1191. doi: 10.1136/gut.2009.205625

Xue, L., Aihara, E., Wang, T. C., and Montrose, M. H. (2011). Trefoil factor 2 requires Na/H exchanger 2 activity to enhance mouse gastric epithelial repair. *J. Biol. Chem.* 286, 38375–38382. doi: 10.1074/jbc.m111.268219

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 14 August 2020
doi: 10.3389/fphys.2020.00954

Fine-Tuning of Energy Levels Regulates *SUC2* via a SNF1-Dependent Feedback Loop

Sebastian Persson [1,2], ***Niek Welkenhuysen*** [1,2†], ***Sviatlana Shashkova*** [3†] ***and Marija Cvijovic*** [1,2*]

[1] Department of Mathematical Sciences, University of Gothenburg, Gothenburg, Sweden, [2] Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden, [3] Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

OPEN ACCESS

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
Matteo Barberis,
University of Surrey, United Kingdom
Jie Zhang,
Technical University of Denmark, Denmark

****Correspondence:***
Marija Cvijovic
marija.cvijovic@chalmers.se

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *30 April 2020*
Accepted: *15 July 2020*
Published: *14 August 2020*

Citation:
Persson S, Welkenhuysen N, Shashkova S and Cvijovic M (2020) Fine-Tuning of Energy Levels Regulates SUC2 via a SNF1-Dependent Feedback Loop. Front. Physiol. 11:954. doi: 10.3389/fphys.2020.00954

Nutrient sensing pathways are playing an important role in cellular response to different energy levels. In budding yeast, *Saccharomyces cerevisiae*, the sucrose non-fermenting protein kinase complex SNF1 is a master regulator of energy homeostasis. It is affected by multiple inputs, among which energy levels is the most prominent. Cells which are exposed to a switch in carbon source availability display a change in the gene expression machinery. It has been shown that the magnitude of the change varies from cell to cell. In a glucose rich environment Snf1/Mig1 pathway represses the expression of its downstream target, such as *SUC2*. However, upon glucose depletion SNF1 is activated which leads to an increase in *SUC2* expression. Our single cell experiments indicate that upon starvation, gene expression pattern of *SUC2* shows rapid increase followed by a decrease to initial state with high cell-to-cell variability. The mechanism behind this behavior is currently unknown. In this work we study the long-term behavior of the Snf1/Mig1 pathway upon glucose starvation with a microfluidics and non-linear mixed effect modeling approach. We show a negative feedback mechanism, involving Snf1 and Reg1, which reduces *SUC2* expression after the initial strong activation. Snf1 kinase activity plays a key role in this feedback mechanism. Our systems biology approach proposes a negative feedback mechanism that works through the SNF1 complex and is controlled by energy levels. We further show that Reg1 likely is involved in the negative feedback mechanism.

Keywords: SNF1, feedback, single-cell, nutrient signaling, dynamic modeling, NLME, STS

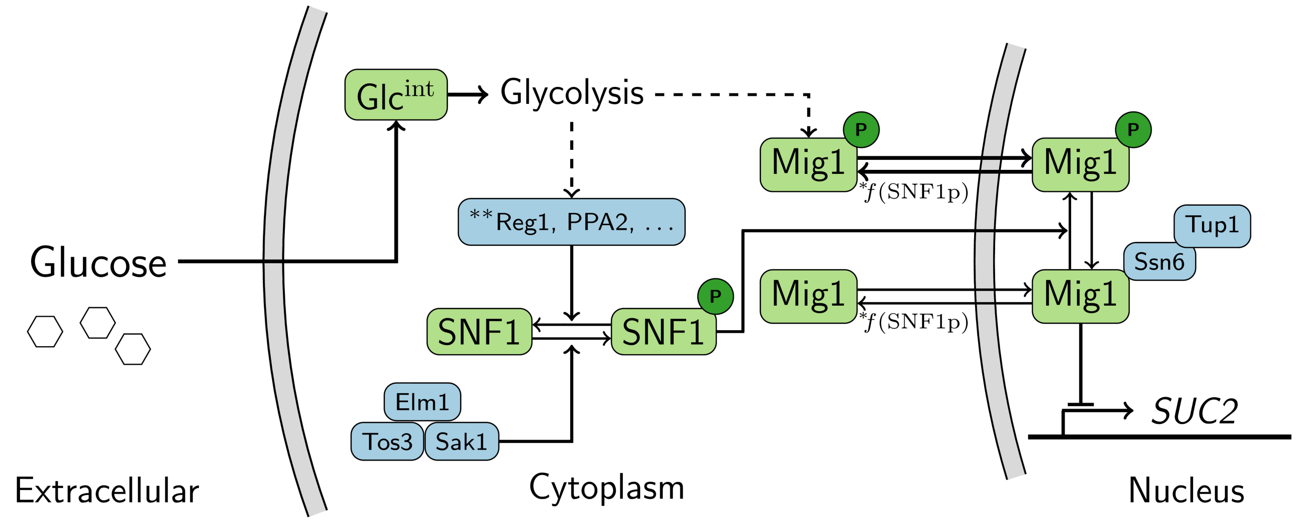
1. INTRODUCTION

Nutrients play a key role in cell survival and well-being by serving as energy sources, cellular building blocks and as triggers for a multitude of signaling pathways. A number of nutrient-controlled signaling pathways has been extensively studied and the crosstalk between them has been elucidated (Shashkova et al., 2015). In the budding yeast, *Saccharomyces cerevisiae*, nutrient controlled pathways distinguish between preferred and alternative nitrogen and carbon sources, and alters the homeostasis to adjust to the extracellular conditions. The Snf1 protein kinase, the yeast orthologue of the mammalian AMP-activated protein kinase (AMPK), regulates energy balance and plays the main role in yeast adaptation to glucose limitation (Carlson et al., 1981; Celenza and Carlson, 1986) via controlling genes required for utilization of non-glucose carbon sources (Treitel et al., 1998). It works in a complex, named SNF1, which is composed of the catalytic subunit Snf4 and one of the three alternative stabilizing subunits Gal83, Sip1, or Sip2 (Jiang and Carlson, 1996; Schmidt and McCartney, 2000). Under glucose depleted conditions, AMPK/Snf1 is activated by three upstream kinases, Tos3, Elm1 and Sak1 (Hong et al., 2003; Nath et al., 2003; Elbing et al., 2006; Rubenstein et al., 2008), which leads to phosphorylation of various transcription factors to facilitate cellular response (Ghillebert et al., 2011). Glucose presence makes the Snf1 activation loop accessible for protein phosphatases (Rubenstein et al., 2008). The Glc7-Reg1 phosphatase is the main negative regulator of Snf1, with a contribution of Sit4 and Ptc1 phosphatases (Rubenstein et al., 2008; Ruiz et al., 2012). At the same time, the assembly and functionality of the Glc7-Reg1 phosphatases depends on Snf1 during poor glucose conditions; active Snf1 phosphorylates Reg1 which prevents its association with the Glc7 subunit to form a functional phosphatase (Sanz et al., 2000). The activation of the SNF1 complex has been shown to correlate with a high ADP/ATP ratio (Rubenstein et al., 2008; Chandrashekarappa et al., 2013). Furthermore, ADP binds to the regulatory subunit of the SNF1 complex resulting in protection of Thr210 from dephosphorylation (Mayer et al., 2011). Overall, this suggests that the SNF1 complex is regulated by intracellular energy levels.

Genes essential for metabolism of maltose (*MAL*), galactose (*GAL*), and sucrose (*SUC2*) are regulated by the transcriptional repressor Mig1, where *SUC2* is one of the most studied (Lutfiyya et al., 1998; Carlson, 1999). Expression of Mig1 target genes is released upon glucose limitation, when Mig1 becomes phosphorylated by Snf1 and relocates to the cytoplasm. In glucose-rich extracellular conditions, the Reg1-Glc7 phosphatase dephosphorylates Mig1 in a glucose-dependent manner, however, another glucose-independent mechanism has been reported to participate in Mig1 dephopshorylation (Shashkova et al., 2017). When dephopshorylated, Mig1 relocates to the nucleus (Treitel and Carlson, 1995; Wu and Trumbly, 1998) where it recruits the Ssn6-Tup1 global co-repressor complex to repress genes (Lutfiyya et al., 1998; Smith et al., 1999; Ahuatzi et al., 2007).

Despite the fact the Snf1/Mig1 pathway (**Figure 1**) has been extensively studied, little is known about how its response to glucose starvation is maintained over time. Thus, it is still unclear how the gene expression response to altered energy levels is fine-tuned, for example, the dynamic regulation of the pathway target genes in the long run. Also, the majority of work on the Snf1/Mig1 pathway has been performed on cell cultures representing the average behavior of the population. When the aim is a cellular level mechanistic understanding of a pathway, such as the Snf1/Mig1 pathway, a single-cell level model is advantageous compared to a population average model. This is due to the fact that population based models do not account for intrinsic heterogeneity within the population (Shashkova and Leake, 2017). Hence, a population based model might overlook dynamic single-cell features, such as oscillations (Cohen-Saidon et al., 2009).

Systems biology approaches exploiting single-cell techniques to study the cellular mechanism of the Snf1/Mig1 pathway (Bendrioua et al., 2014; Almquist et al., 2015; Welkenhuysen et al., 2017) have previously been employed. However, these have focused on the short-term localization of Mig1 as measurement for the Snf1/Mig1 pathway activity. Long time-lapse observation



$^{*}f$(SNF1p) = Transport out of nucleus depends on (is a function) of SNF1p activity

** Besides Reg1-Glc7 and PPA2 there are other components (...) regulating SNF1

FIGURE 1 | Schematic representation of the Snf1/Mig1 pathway. Snf1 is constitutively phosphorylated by the upstream kinases. When preferred carbon-sources are limited Snf1 phosphorylates Mig1, which results in the cytoplasmic localization of Mig1. When preferred carbon sources are available Snf1 and Mig1 are dephosphorylated through an unknown mechanism originating from glycolysis. Mig1 localizes to the nucleus and there it represses genes which are required for the utilization of non-fermentable carbon-sources. Dashed arrows represent signals that are transmitted via unknown mechanisms.

of pathways with microfluidics device provide high-quality observation rich data. In combination with mathematical modeling this data can be used to understand complex behavior of cellular pathways. The abundance of species in Snf1/Mig1 pathway (Shashkova et al., 2015) together with previous modeling studies (Almquist et al., 2015), suggest that the cell-to-cell variability in Snf1/Mig1 pathway mostly arises from differences in protein levels (extrinsic noise). Therefore, a suitable modeling framework to study single-cell activity of the Snf1/Mig1 pathway is non-linear mixed effects modeling (NLME) (Almquist et al., 2015; Llamosi et al., 2016).

In this work we study the long-term behavior of the Snf1/Mig1 pathway upon glucose starvation with a microfluidics and NLME modeling approach. We employ a single-cell microfluidics technique to capture the long-term behavior of a large set of cells after the nutrients conditions in the cell environment have changed. We further examined activity levels of the invertase enzyme as a readout of the *SUC2* expression to investigate the effect of Snf1 and Reg1 on target genes. Moreover, we utilize a standard fluorescence microscopy approach on cells subjected to a pharmacological inhibitor to show the crucial role of the Snf1 kinase activity in regulation of Mig1 localization that defines its function on target promoters. We show that a feedback mechanism reduces *SUC2* expression after the initial strong activation. Our modeling approach proposes a negative feedback mechanism that works through the SNF1 complex and is controlled by energy levels. We further show that Reg1 is involved in the negative feedback mechanism.

2. MATERIALS AND METHODS

2.1. Yeast Strains and Growth Conditions

Standard YPD (10 g/l yeast extract, 20 g/l bacto-peptone, glucose according to experimental needs) and YNB [1.7 g/l yeast nitrogen base without amino acids, without $(NH_4)_2SO_4$, 5 g/l $(NH_4)_2SO_4$, supplemented with glucose and amino acids according to nutritional requirements] media were used for yeast cells growth and transformants selection.

To delete the *SNF1* gene, the *LEU2* fragment from YDp-L plasmid (Berben et al., 1991) flanked on its 5′- and 3′-termini with 50 bp up- and downstream of *SNF1*, respectively, was amplified by PCR. Strain YSH2348 was transformed directly with the PCR reaction mix by standard lithium acetate protocol (Gietz and Schiestl, 2007) and placed on YNB leucin-deficient agar plates supplemented with 4% glucose. Successful transformants were verified by confirmation PCR.

2.2. Invertase Assay

Pre-grown cells were inoculated into 50 ml of fresh YPD medium with 4% glucose and grown until mid-log phase. A half of a culture then was harvested by rapid centrifugation and freezing in liquid nitrogen, the other half was washed with water, suspended in a fresh YPD medium with 0.2% glucose, incubated at 30°C, 180 rpm for following 4 h and harvested as above. Yeast cells were mechanically disrupted by glass beads in crude extraction buffer [50 mM imidazole, 100 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1× protease inhibitor cocktail (Roche)], and the lysates we obtained by collecting supernatants after centrifugation for 10 min at 10,000 rpm. The total amount of protein was quantified by RC DC protein quantification kit (Bio-Rad). Protein extracts were mixed with acetate buffer (0.3 M CH_3COOH, 0.2 M CH_3COOK), and the reaction was initiated by adding 500 mM sucrose solution (in 0.1 M CH_3COOK). After 10 min the reaction mix was added to 0.1 M KPO_4 pH 6.5 containing peroxidase, glucose oxidase, and O-dianisidine, and incubated at 30°C water bath for 15 min. The reaction was stopped by adding 6M HCl, and the OD at 540 nm was measured. One unit of invertase activity is the amount of enzyme that produces 1 nmol of glucose per minute at pH 6.5 at 30°C.

2.3. Epifluorescence Microscopy

Pre-grown cells carrying *pSNF1-TAP* or its’ analog-sensitive version, *pSNF1-I132G-TAP*, were cultivated in YNB medium with uracil-deficient amino acid supplement with 0.2% glucose for 1 h. To block Snf1 phosphorylation, 2 μM ATP-competitive kinase inhibitor, 1NM-PP1 (Cayman), was added to the cell cultures for 5 min at room temperature. For the wide field fluorescence microscopy cells were imaged using an ApoTome camera and a Zeiss Axiovert 200M microscope (Carl Zeiss MicroImaging). Fluorescence images were acquired by using separate filter sets 38HE and 43HE for GFP and mCherry excitation, respectively.

2.4. Time-Lapse Microscopy and Cell Tracking

The yeast cells (W303, HXK1p-Citrine-ACT1t, **Table S3**) were grown overnight and injected with a syringe in a two-channel Y-formed microfluidics poly-dimethylsiloxane (PDMS) system and allowed to sediment in the main channel. For the switch fresh CSM media was supplied to the cells through the other channel. The experimental setup is further described in Welkenhuysen et al. (2018). Imaging was performed on a Leica DMi8 inverted fluorescence microscope (Leica microsystems). The microscope was equipped with a HCX PL APO 40×/1.30 oil objective (Leica microsystems), Lumencor SOLA SE (Lumencor) led light and Leica DFC9000 GT sCMOS camera (Leica microsystems). Cell growth was recorded at 1 frame in bright-field at 20 ms exposure, and YFP was observed with an excitation: 500/20, dichroic: 515 and emission: 535/30 filtercube at 150 ms exposure every 5 min. Analysis of fluorescence intensity was performed with the ImageJ distribution FIJI (Schindelin et al., 2012).

2.5. Feedback Cascade Model

The simple feedback cascade model (**Figure 3A**) is based on ODEs with the rate-equations formulated using mass action and Hill-kinetics (Equation 1). The model has three components: Snf1/Mig1 pathway inhibitory activity (*SNF*1*pat*), *SUC2* and a potential feedback cascade (X). The model investigates if a partial recovery in intracellular energy levels, that results in a recovery in *SNF*1*pat* activity (meaning the inactivation of Snf1 and activation of Mig1) (Mayer et al., 2011), might explain the observed *SUC2* expression. This is achieved by including the likely recovery in intracellular energy levels, which is an effect of genes activated upon a glucose drop and the decreased activity of *SNF1pat* (Hedbacker and Carlson, 2008), via the production of *X* (Equation 1c). Furthermore, as a partial recovery in energy levels should result in increased activity of the energy regulated *SNF1pat*, *X* promotes production of *SNF1pat* (Equation 1a). That is, *X* partially restores *SNF1pat* activity, which is a measure of the intracellular energy levels during high glucose conditions (i.e., high intracellular energy level).

$$
\begin{aligned}
\frac{\mathrm{d}SNF1pat(t)}{\mathrm{d}t} &= k_{\mathrm{glc\_ex}}(t) - k_2 SNF1pat(t) + k_3 X(t-\tau_{\mathrm{f}}) && \text{(1a)}\\
\frac{\mathrm{d}SUC2(t)}{\mathrm{d}t} &= \frac{k_4}{k_5 + SNF1pat(t-\tau_{\mathrm{m}})} - k_6 SUC2(t) && \text{(1b)}\\
\frac{\mathrm{d}X(t)}{\mathrm{d}t} &= \mathcal{H}(t)\frac{k_7}{k_8 + SNF1pat(t)} - k_9 X(t) && \text{(1c)}\\
SNF1pat(t) &= SNF1pat_{t_0} \quad \forall t \in [t_0 - \tau_{\mathrm{m}}, t_0] && \text{(1d)}\\
SUC2(t) &= SUC2_{t_0}, \quad t = t_0 && \text{(1e)}\\
X(t) &= 0, \quad \forall t \in [t_0 - \tau_{\mathrm{f}}, t_0] && \text{(1f)}
\end{aligned}
$$

Besides standard model building procedures, three challenges had to be addressed when constructing the model. Firstly, the model aims to describe the *SUC2* expression. However, the *SUC2* expression is measured via matured YFP. Secondly, the feedback affecting the SNF1 pathway is likely a cascade of events. The fluorescence maturation and feedback cascade can be represented by adding multiple states to the model. However, to keep the model identifiable, time-delays τ_i were used instead (**Figure 3A**). As the data shows a variation in feedback response-time (**Figure 2B**), τ_{f} was estimated from the data. As the YFP maturation time has been reported to be 39 ± 7 min (Gordon et al., 2007), and the data shows an increase in YFP at 32 min (**Figure 2B**), τ_{m} was fixated at 32 min. Thirdly, at time zero the model is at rich glucose conditions. At rich glucose conditions, the SNF1 pathway, and thus the model should be at steady state. Hence, the rate equations were forced to be zero at time zero, which yielded expressions for initial values, as functions of rate constants, that ensures an initial steady-state (Equation 2b). The glucose downshift which breaks the steady state, was included in the model by reducing glucose in-signal and activating the feedback mechanism close to time zero (Equation 2a). It should be noted that *X* initially in high glucose, is modeled to not interact with *SNF1pat* and thus has zero activity (Equations 1c and 2a). This is because the feedback is modeled to be dependent on the energy saving processes that are activated upon an external glucose drop (Hedbacker and Carlson, 2008). In addition to the information presented here, a more detailed description of each state and accompanying rate equations is presented in **Table S1**.

$$
k_{\mathrm{glc\_ex}}(t) = \begin{cases} k_1 & t < 0.0483 \\ k_1/40 & t \geq 0.0483 \end{cases}, \quad \mathcal{H}(t) = \begin{cases} 0 & t < 0.0483 \\ 1 & t \geq 0.0483 \end{cases} \tag{2a}
$$

$$
SNF1pat_{t_0} = k_1/k_2, \qquad SUC2_{t_0} = \frac{k_4}{\left(k_5 + SNF1pat_{t_0}\right)k_6} \tag{2b}
$$

When fitting the model to the *SUC2* data, all unknown parameters except two (k_5 and k_8) were estimated with a full random effects covariance matrix $\boldsymbol{\Omega}$. These two parameters were estimated without random effects to reduce the size of $\boldsymbol{\Omega}$, and consequently keeping the standard errors of the estimated parameters acceptable (**Table S4**). It should also be noted that fixing k_5 and k_8 does not, unrealistically, removes cell-to-cell variability in the reactions they govern. This is because k_4 and k_7 are assumed to vary between cells (Equations 1b and 1c). The unidentifiable Hill-coefficients (Equations 1b and 1c) where fixed to the smallest integer that produced a good fit (n = 1). In addition to the information presented here, inference details for each parameter is presented in **Table S1**.

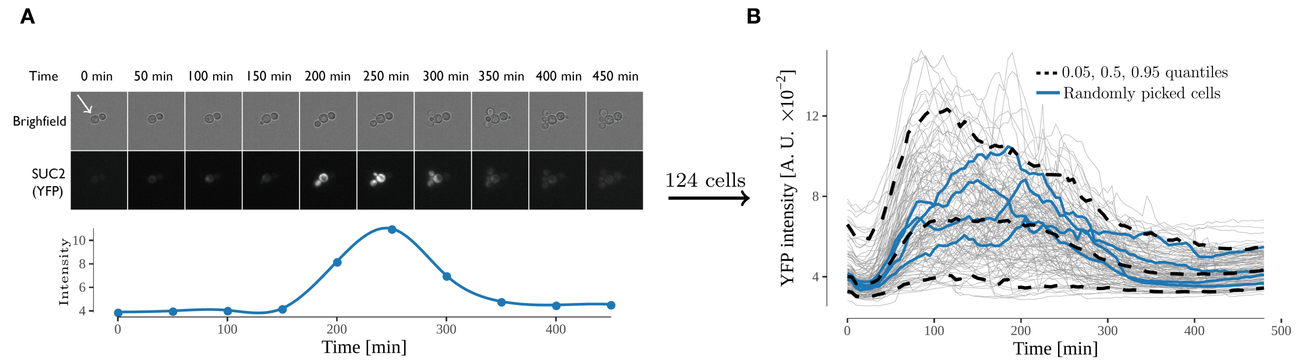


FIGURE 2 | Expression from the *SUC2* promotor over time, expressed as fluorescence intensity of YFP divided by area. **(A)** time-lapse images of cells exposed to shift from high (4%) to low (0.1%) glucose. Both bright field and yellow fluorescence images are displayed. Below is the intensity over time of one cell (annotated by white arrow). The fluorescence intensity was observed for 480 min with an interval of 5 min between every image acquisition. **(B)** Fluorescence intensity over time of all analyzed cells. Each gray line represents the trace of one single cell, and the 0.05, 0.5, and 0.95 quantiles of the cells are represented with a dashed black line. Four randomly selected cells are represented with a blue line. 124 cells were analyzed in the experiment.

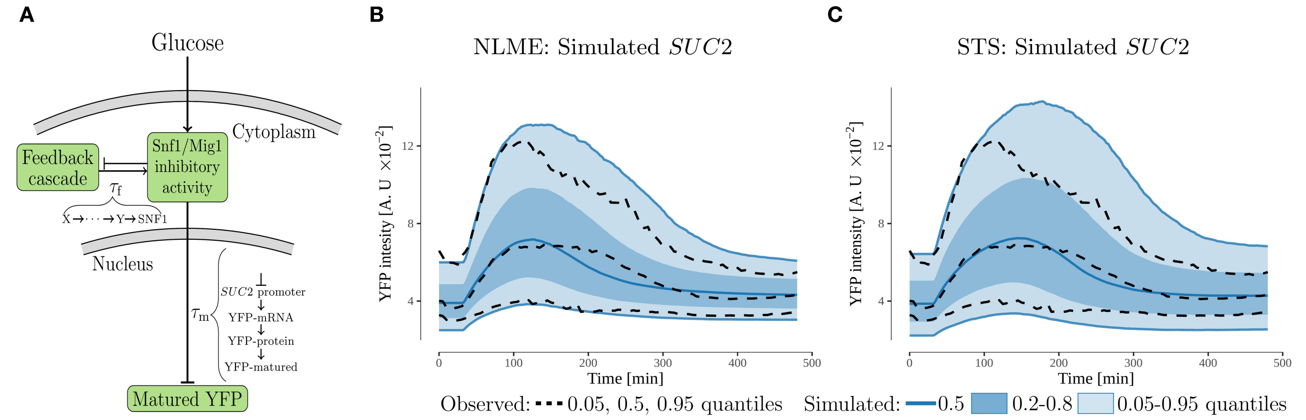


FIGURE 3 | The feedback cascade model. **(A)** Schematic representation. The model proposes that the *SUC2* expression decreases during glucose starvation due to a feedback acting on the inhibitory activity of the Snf1/Mig1 pathway. A potential source of this feedback is partial recovery in the intracellular energy levels. The feedback-cascade, and maturation time of the promoter$_{SUC2}$-YFP (measured output), were represented by time-delays, τ_i. **(B,C)** Simulated and observed population behavior of the SUC2-intensity (note same scale on y-axis). The simulated data was generated by simulating 10,000 cells using the estimated parameter distributions for the NLME and STS approaches.

2.6. Feedback Mediated Model

The feedback mediated model (**Figure 4D**) accounts for phosphorylated and dephosphorylated forms of nuclear Mig1 as well as for observed movement of Mig1 out of the nucleus upon a glucose downshift. The model consists of five states, phosphorylated (active) SNF1-complex ($SNF1p$), nuclear Mig1 ($Mig1$), phosphorylated nuclear Mig1 ($Mig1p$), $SUC2$ and a feedback mediating component (Y). All rate-equations, excluding Mig1 nuclear transport, were constructed using mass action and Hill kinetics (Equation 3). A sigmoid function, $\sigma(SNF1p)$, was implemented and parameterized to match observed Mig1 behavior (Equation 4b, for parametrization details see **Figure S3**), so that Mig1 stays in the nucleus during low SNF1 activity (**Figure 4A**), and that roughly as much Mig1 moves out of the nucleus when the external glucose is reduced from 4 to 1.5%, as when it is reduced from 4 to 0.1% (Bendrioua et al., 2014). The model investigates if a recovery in intracellular energy, resulting in a decrease in $SNF1p$ activity (Mayer et al., 2011), can explain the observed *SUC2* expression. This is achieved by including the likely recovery in intracellular energy levels, which is an effect of genes activated upon glucose starvation and the increased SNF1p-activity (which results in increased *SUC2* expression) (Hedbacker and Carlson, 2008), via the production of Y (Equation 3e). Furthermore, as a partial recovery in energy levels should result in reduced SNF1-complex activity and hence *SUC2* activity, Y inhibits formation of $SNF1p$ (Equation 3a).

$$\frac{\mathrm{d}SNF1p(t)}{\mathrm{d}t} = k_1 - k_{\mathrm{ex\_glc}}(t)\mathcal{A}(t)SNF1p(t) - k_{10}Y(t)SNF1p(t) \tag{3a}$$

$$\frac{\mathrm{d}Mig1(t)}{\mathrm{d}t} = k_2 - k_3 SNF1p(t)Mig1(t) + k_4 Mig1p(t) - k_5\Big(1 + \sigma\big(SNF1p(t)\big)\Big)Mig1(t) \tag{3b}$$

$$\frac{\mathrm{d}Mig1p(t)}{\mathrm{d}t} = k_3 SNF1p(t)Mig1(t) - k_4 Mig1p(t) - k_5\Big(1 + \sigma\big(SNF1p(t)\big)\Big)Mig1p(t) \tag{3c}$$

$$\frac{\mathrm{d}SUC2(t)}{\mathrm{d}t} = \frac{k_6}{K + Mig1(t - \tau_{\mathrm{m}})} - k_7 SUC2(t) \tag{3d}$$

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = k_8\big(SUC2(t) - SUC2_{t_0}\big) - k_9 Y(t) \tag{3e}$$

$$SNF1p(t) = 0, \quad t = t_0 \tag{3f}$$

$$Mig1(t) = Mig1_{t_0} \quad \forall t \in [t_0 - \tau_{\mathrm{m}}, t_0] \tag{3g}$$

$$Mig1p(t) = 0, \quad t = t_0 \tag{3h}$$

$$SUC2(t) = SUC2_{t_0}, \quad t = t_0 \tag{3i}$$

$$Y(t) = 0, \quad t = t_0 \tag{3j}$$

The initial values for phosphorylated $Mig1p$ and $SNF1$ were set to 0 as the model is in rich glucose conditions at time zero. By the same arguments as for the simple feedback model, the initial value for the feedback component (Y) was fixed to 0, the YFP-maturation was represented by a time-delay, and a steady state which is broken by a reduction in the external glucose signal was enforced at time zero (Equation 4a). To ensure an initial steady state for $SNF1p$, it was noted that the external glucose signal should be heavily amplified ($\mathcal{A}$) during high glucose conditions (Equation 4a and **Table S2**), as multiple processes negatively regulates $SNF1p$ during high glucose conditions (Ruiz et al., 2011, 2013; Zhang et al., 2011; McCartney et al., 2016). Furthermore, by assuming a steady state expression for the initial values of $Mig1$ and $SUC2$ were obtained as a function of rate parameters (Equation 4c), and an expression of $k_{\mathrm{glc\_ext}}(t)$ was obtained as a function of k_1 (Equation 4a). Lastly, as the simple model suggested a delayed feedback, this was added in the model by making the feedback dependent on the YFP activity of the Snf1/Mig1 regulated *SUC2* gene. In addition to the information presented here, a more detailed description

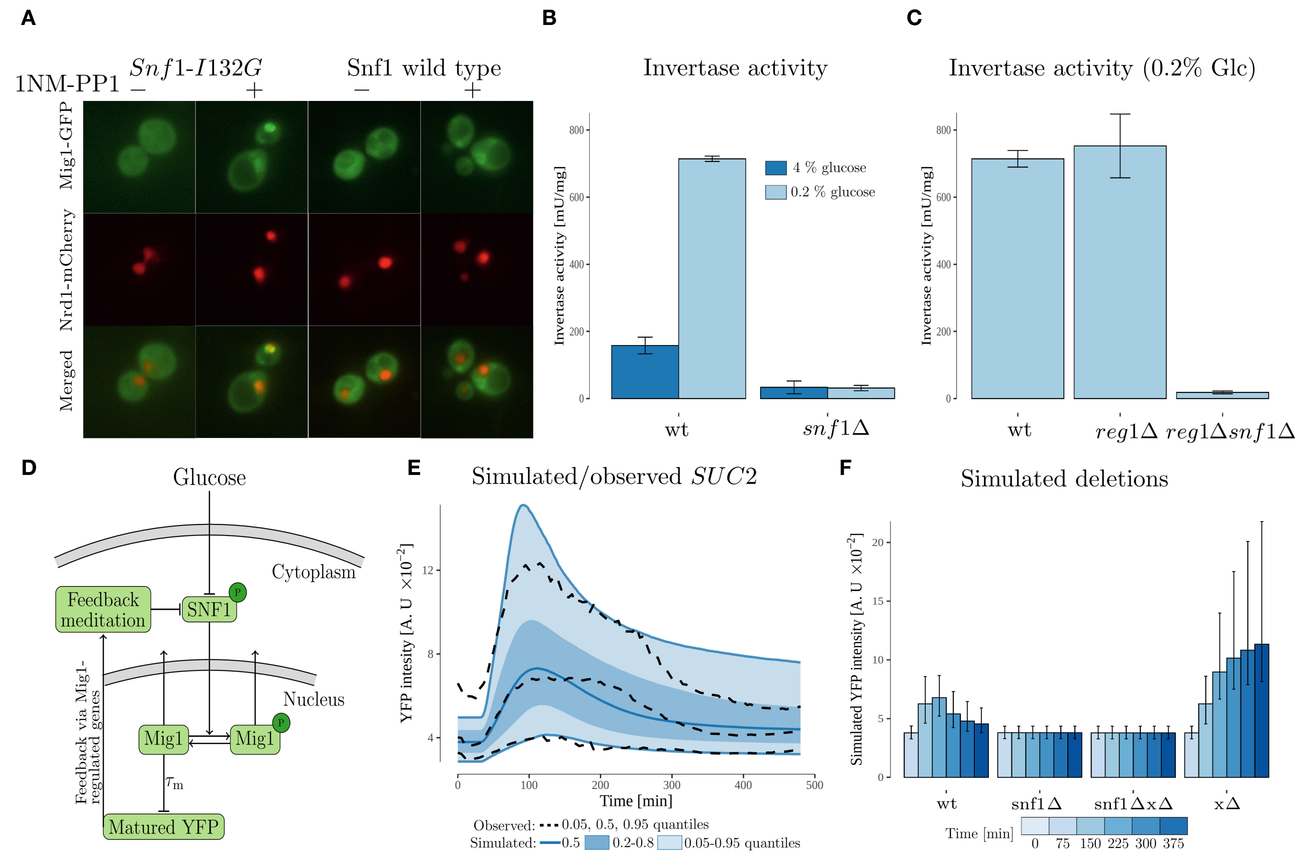


FIGURE 4 | Experimental data and modeling suggests *SUC2* regulating feedback via SNF1 complex. **(A)** Cells with genomically integrated Mig1-GFP and Nrd1-mCherry fusions and expressing the wild type Snf1 or its analog-sensitive version, *Snf1-I132G*, were incubated in low (0.2%) glucose conditions with subsequent addition of 1NM-PP1 for 5 min prior to imaging. Images were enhanced using FiJi software for representation. **(B)** A read-out of invertase activity measured in cell lysates obtained from cultures pre-grown in 4% glucose medium, then incubated in 0.2% glucose conditions for 4 h. Error bars represent standard deviation. **(C)** A read-out of invertase activity measured in cell lysates obtained from cultures pre-grown in 4% glucose medium, then incubated in 0.2% glucose conditions for 4 h. Error bars represent standard deviation. **(D)** Feedback mediated model. The model, which is an extension of the cascade model (**Figure 3**), proposes that the *SUC2* expression decreases upon long term glucose starvation due to a feedback acting on the phosphorylated SNF1-complex (*SNF1p*). The maturation time of the promoter$_{SUC2}$-YFP (measured output) was represented by a time-delay. The Mig1 nuclear export was represented by a *SNF1p* dependent sigmoid function. **(E)** Simulated and observed *SUC2* expression. The simulated data was obtained by simulating 10,000 cells from the feedback mediated model. **(F)** Simulated deletions obtained by simulating 10,000 cells for each case from the feedback mediated model. Simulated intensity was sampled at discrete time-points. Error bars represent the 0.20 and 0.80 quantiles within the simulated single cell population.

of each state and accompanying rate equations is presented in **Table S2**.

$$k_{\text{ex\_glc}}(t)\mathcal{A}(t)SNF1p(t) = \begin{cases} kGlc^{\text{ext}}\underbrace{\mathcal{A}(t)SNF1p(t)}_{\approx 1} \approx kGlc^{\text{ext}} \overset{\text{steady state}}{=} k_1, & t < 0.0483 \\ k\frac{Glc^{ext}}{40}\underbrace{\mathcal{A}(t)}_{=1}SNF1p(t) = \frac{k_1}{40}SNF1p(t), & t \geq 0.0483 \end{cases} \tag{4a}$$

$$\sigma\big(SNF1p(t)\big) = \frac{1}{1+\exp\Big(-3\big(SNF1p(t)-4.5/3\big)\Big)} \tag{4b}$$

$$Mig1_{t0} = \frac{k_2}{k_5\Big(1+\sigma\big(SNF1p(t_0)\big)\Big)}$$

$$SUC2_{t_0} = \frac{k_6}{(K+Mig1_{t0})k_7} \tag{4c}$$

When fitting the model to the *SUC2* expression data all unknown parameters, except one which was fixed ($K = 0.1$), were estimated with a full random effects covariance matrix $\boldsymbol{\Omega}$. This approach reduces the size of $\boldsymbol{\Omega}$ and consequently ensures acceptable standard errors of the estimated parameters (**Table S5**). The unidentifiable Hill-coefficient (Equation 3d) was fixed to the smallest integer that produced a good fit ($n = 1$). In addition to the information presented here, inference details for each parameter is presented in **Table S2**.

It should be noted that the reaction-scheme in **Figure 4D** can also be modeled by an approach that requires less assumptions, and is more consistent with the feedback cascade model (Equation 1). This by replacing $k_{\text{ex\_glc}}ASNF1p$ with; $-k_{ex\_glc} -$

k_{11} *SNF*1*p*, a steady state argument will result in the same expression for k_{ex_glc} as in Equation (4a) ($k_{\text{ex\_glc}} = k_1, t < 0.0483$ and $k_{\text{ex\_glc}} = k_1/40, t \geq 0.0483$). Overall, this approach also produces a good fit (**Figure S4A**), with the same dynamic characteristics as in the feedback mediated model, e.g., Mig1 partially relocates back into the nucleus upon long term glucose starvation (**Figure S4B**). However, this approach also results in a non-identifiable model, since by adding the k_{11} rate-parameter an additional 12 parameters have to be estimated due to the random effects covariance matrix. Hence, we used the model presented in Equation (3) as it is identifiable, is biologically justifiable (**Table S2**), captures key biological features, such as having practically zero *SNF*1*p* activity during high glucose (McCartney and Schmidt, 2001), and also keeps the dynamics of the approach with less assumptions.

Furthermore, it should be noted that the SNF1 related state variables in the two models correspond to different entities. In the feedback cascade model (**Figure 3A**), *SNF*1*pat* represents the inhibitory activity of the Snf1/Mig1 system. The inhibitory, instead of the activating activity, of the Snf1/Mig1 pathway was implemented in the model as the Snf1/Mig1 pathway generally is reported to inhibit the activity of the *SUC2* promoter (Broach, 2012). As the feedback mediated model (**Figure 4D**) is more detailed regarding the Snf1/Mig1 pathway, the SNF1 related state variable *SNF*1*p* represents the activity of phosphorylated SNF1-complex. The reason for using *SNF*1*p*, instead of *SNF*1 (non-phosphorylated complex), is 2-fold. Firstly, our data show how *SNF*1*p* affects Mig1 nuclear export (**Figure 4A**). Secondly, Mig1 is phosphorylated in a *SNF*1*p*-dependent manner (Shashkova et al., 2017). However, it is not known how SNF1 affects dephosphorylation of *Mig*1*p*, hence a term $Mig1p \cdot SNF1$ would not be justifiable to include in the model. As the SNF1 related state variables correspond to different entities, they reflect the amount of intracellular energy levels, which SNF1 senses via the ATP/ADP ratio (section 3.2), differently. More precisely, high intracellular energy levels is represented by high *SNF*1*pat* and low *SNF*1*p*, with the conversely holding for low intracellular energy levels.

2.7. Parameter Estimation

In an ODE-model where the unknown model parameters are allowed to vary between cells, the dynamics of cells $i = 1, \ldots, n$ are given by

$$\frac{\mathrm{d}\mathbf{x}_i(t_j)}{\mathrm{d}t} = \mathbf{g}\big(\mathbf{x}_i, \mathbf{k}_i, \mathbf{u}(t_j), t_j\big), \quad \mathbf{x}_i(t_0) = \mathbf{x}_{i0}, \tag{5}$$

where $\mathbf{g}$ is the reaction kinetics governed by the value of the model states $\mathbf{x_i}$, kinetic parameters $\mathbf{k}_i$ and potential input functions $\mathbf{u}(t)$. In NLME and the standard two stage (STS) approach, the goal is to estimate the unknown model parameters $\mathbf{b}_i = (\mathbf{k}_i, \mathbf{x}_{i0})$ and their underlying parameter distribution using observed data $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$. In the course of this work, the models (**Figures 3A**, **4D**), were related to the observed *SUC2* expression data (**Figure 2B**) via, as in a previous study with similar data (Almquist et al., 2015), an additive error model

$$y_i(t_j) = y_{ij} = \underbrace{\hat{y}_{ij}(\mathbf{b}_i)}_{\text{simulated SUC2}} + e_{ij}, \quad e_{ij} \sim \mathcal{N}(0, s^2), \tag{6}$$

where s^2 is the variance of the measurement error.

2.7.1. Non-linear Mixed Effect Approach

In the non-linear mixed effect (NLME) approach, the unknown parameters are split into a fixed effect $\mathbf{b}$, and an individual random effect $\boldsymbol{\eta}_i$. Here, the individual parameters were assumed to be lognormally distributed; $\mathbf{b}_i = \mathbf{b}e^{\boldsymbol{\eta}_i}$, $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. Lognormal parameters were assumed instead of normally distributed parameters by two reasons. Firstly, lognormality ensures positive parameters. Secondly, rate constants are generally products of different factors (Atkins et al., 2013). Thus, a lognormal distribution is more suitable than a normal distribution (Limpert et al., 2001).

The fixed effects, covariance matrix of random effect and the noise parameter, $\boldsymbol{\psi} = (\mathbf{b}, \boldsymbol{\Omega}, s^2)$, are estimated simultaneously in NLME by maximizing a likelihood function

$$\mathcal{L}(\boldsymbol{\psi}|\mathbf{y}) = \prod_{i=1}^{n} \int p(\mathbf{y}_i|\boldsymbol{\eta}_i, \mathbf{b}, s^2) p(\boldsymbol{\eta}_i|\boldsymbol{\Omega}) \mathrm{d}\boldsymbol{\eta}_i. \tag{7}$$

The integral in Equation (7) is generally intractable, and is computationally expensive to approximate (Davidian and Giltinan, 2003). Consequently, other methods than the classical optimization methods must be used to maximize the likelihood. Here the SAEM-algorithm (Kuhn and Lavielle, 2005), via the Monolix software (Lixoft, 2019), was used. The SAEM-algorithm was chosen as it works well for NLME-models (Chan et al., 2011), and it's convergence has been rigorously proven (Delyon et al., 1999). Given estimated population parameters $\hat{\boldsymbol{\psi}} = (\hat{\mathrm{b}}, \hat{\boldsymbol{\Omega}}, \hat{s})$, new cells were simulated by first randomly drawing $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \hat{\boldsymbol{\Omega}})$, and then calculating the parameters corresponding to a simulated cell by $\mathbf{b}_{\text{simulated}} = \hat{\mathbf{b}}e^{\mathbf{a}}$.

To construct diagnostic plots based on individual predictions (IPRED-plots **Figures S1A**, **S2A**), the empirical Bayes estimates (EBS:s) were used; $\mathbf{b}_i^{(\text{EBE})} = \arg\max p(\mathbf{b}_i|\hat{\boldsymbol{\psi}}, \mathbf{y}_i)$ (Davidian and Giltinan, 2003). To construct diagnostic plots of the parameter distributions (Q-Q plots **Figure S1B**), random samples from the conditional distribution $p(\boldsymbol{\psi}_i|\mathbf{y}_i)$ were used instead of the EBE:s, avoiding potential effects of η-shrinkage on model diagnostics (Lavielle and Ribba, 2016).

Identifiability was investigated by calculating the standard errors of the estimated parameters. The standard errors were obtained by inverting the observed Fisher Information Matrix (FIM), which was calculated via a stochastic approximation algorithm implemented in Monolix (Lixoft, 2019). The method of profile likelihood (Raue et al., 2009), is typically superior to asymptotic approaches, such as FIM for investigating model identifiability. However, the computational burden of optimizing and evaluating the likelihood (Equation 7), makes the profile likelihood method too computationally expensive to be used here.

2.7.2. Standard Two Stage Approach

In the standard two stage (STS) approach, $\mathbf{b_i}$ is first estimated for each cell. Given individual estimates $\hat{\mathbf{b}}_i$, the parameter distribution is inferred. The parameters were assumed to follow a lognormal distribution (section 2.7.1).

The individual parameters were estimated by the maximum likelihood method (Llamosi et al., 2016). To avoid a constrained optimization problem, and to facilitate a more efficient optimization (Raue et al., 2013), the parameters were estimated on the log-scale. Given the error model (Equation 6), the log-scale parameters $\mathbf{lb}_i$ are obtained by maximizing the log-likelihood

$$
\begin{aligned}
(\hat{\mathbf{lb}}_i, \hat{s}_i) \;=\; & \underset{\mathbf{lb}_i, s_i}{\arg\max} -\frac{n_i}{2}\ln s_i^2 - \frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}\left(y_{ij} - \hat{y}_{ij}(\mathbf{lb}_i)\right)^2 \\
& - \underbrace{\frac{n_i}{2}\ln 2\pi}_{=\text{constant}} . \qquad (8)
\end{aligned}
$$

To solve the optimization problem (Equation 8), the BOBYQA algorithm (Powell, 2009) was used.

Given estimated individual parameters on the log-scale $\hat{\mathbf{lb}}_i$, simulated cells were obtained by: (1) Noting that as the model parameters are assumed to be log-normal, the log-scale parameters $\mathbf{lb}_i$ become multivariate normal. (2) Via the method of maximum likelihood estimating the parameters of this multivariate normal distribution

$$
\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{lb}_i, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{lb}_i - \hat{\boldsymbol{\mu}})(\mathbf{lb}_i - \hat{\boldsymbol{\mu}})^{\mathrm{T}}. \qquad (9)
$$

(3) generating simulated parameter vectors on the log-scale, that is a simulated cells, by randomly drawing vectors from $\mathbf{lb}_{\text{simulated}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$.

2.7.3. Comparison of STS and NLME Approach

Two parameters, k_5 and k_8, in the feedback cascade model (**Figure 3A** and Equation 1) were assumed to not vary between cells. Thus, they were estimated without random effect using the NLME-framework. There is no obvious, single, way to include parameters without random effect in the STS-approach. Thus, two different approaches were used for the STS-estimation. (i) All parameters were allowed to vary between cells. (ii) All parameters were allowed to vary in a first parameter estimation run. Then in a second run k_5 and k_8 were fixed according to the mean-values obtained in the first run. Approach (ii) yielded better results, as approach (i) to a larger degree overestimated the variability when simulating the population behavior (**Figure S1F**), and is thus used for comparison.

2.7.4. Implementation Details

In the implemented NLME approach, the population parameter $\boldsymbol{\psi}$, EBE:s, samples from $p(\boldsymbol{\psi}_i|\mathbf{y}_i)$ and the standard errors were calculated using Monolix (version 2019R2) (Lixoft, 2019). For the STS approach, the individual parameters (Equation 8) were estimated using the BOBYAQ algorithm in the NLopt (version 0.5.1) library in the Julia (version 1.3.1) programming language (Powell, 2009; Bezanson et al., 2017; Johnson, 2020). To ensure a fair comparison between NLME and STS, the same starting values were used for both approaches. For the cascade feedback model (**Figure 3A**) starting values were obtained by choosing the parameters values which produced the best model fit to the observed mean value using a multiple shooting approach. For the feedback mediated model (**Figure 4D**) this approach was not feasible due to stability issues. Consequently, starting values were obtained by taking the population parameters obtained when running the SAEM algorithm on manually adjusted starting values. To simulate new cells, the DifferentialEquations (version 6.11.0) library in Julia was used for solving the delay-differential equation system that makes out the models (Rackauckas and Nie, 2017). All the calculations were performed on a Dell Latitude with eight cores [Intel(R) Core(TM) i5-8365U CPU @ 1.60 GHz] running on Ubuntu 18.04.4.

The code used to produce all results in this paper can be found on GitHub (https://github.com/cvijoviclab/SUC2_long_term_regulation). Efforts have been made to make the result as reproducible as possible by basing the directory structure on two suggestions (Noble, 2009; Wilson et al., 2017). Given a Unix-based operative system, Monolix (version 2019R2) and Julia (version 1.3.1) the results should be reproducible by running the Run_all-script. More details about reproducing the results can be found on GitHub.

3. RESULTS

3.1. Long-Term Observation of *SUC2* Promoter Expression Reveals Regulation of Promoter Activity After Initial Activation

For a pathway to react appropriately to a stimuli it first needs to be activated, and thereafter the activation needs to be regulated according to the strength of the stimuli. For the Snf1/Mig1 pathway, the current understanding can only result in the monotonic behavior of the activation (Welkenhuysen et al., 2017). That is, if a cell is presented with glucose depletion Mig1 leaves the nucleus and remains there until glucose is available again in the cellular environment. However, recent long-term continuous observation of Snf1/Mig1 pathway has shown pulsatile behavior over an extended period of glucose availability (Dalal et al., 2014; Lin et al., 2015). To elucidate the mechanism behind the long-term adjustment of the Snf1/Mig1 pathway activity upon glucose depletion yeast cells grown on 4% glucose were exposed to a shift to 0.1% in a microfluidic device. These cells contained a construct carrying YFP behind a *SUC2* promoter, allowing for measurement of the *SUC2* expression used to calibrate the models (**Figure 2A**). The *SUC2* gene encodes for two types of invertase, a secreted glycosylated form and an intracellular, non-glycosylated form. The former is regulated by the Snf1-pathway while the latter is produced constitutively in small amounts compared to the glycosylated form (Carlson and Botstein, 1982). Therefore, the non-glycosylated form was not considered in the developed models. The fluorescence intensity was observed for 480 min with an interval of 5 min, yielding a data-set with a rich amount of data points and cells (124 cells with

97 data points each). Initially the level of fluorescent intensity increased 212% on average (**Figure 2B**), thereafter between time points 180 and 210 min the signal started to decline. The decline of signal occurs when the production of new protein is smaller then the turn-over of the YFP protein through breakdown and bleaching. Hence, the decline implies that a decrease of *SUC2* promoter activity has taken place. This suggests that after an initial activation time a negative feedback takes place in the nutrient signaling network which reduces the expression of the *SUC2* promoter.

3.2. Modeling Suggests Delayed Negative Feedback Due to a Partial Recovery in Intracellular Energy Levels After Initial Activation

The single-cell *SUC2* expression data suggests the existence of a negative feedback, which upon long-term glucose starvation reduces the *SUC2* expression (**Figure 2B**). A possible source of this feedback is a partial recovery in the intracellular energy levels, that is an increase in cellular metabolism which results in an increase of the ATP/ADP ratio which further regulates the SNF1 complex. More specifically, during low intracellular energy levels the SNF1-complex is bound by ADP and protected (sterically) against deposphorylation, while the phosphorylation of the complex occurs continuously (Rubenstein et al., 2008; Chandrashekarappa et al., 2013). Conversely, during rich intracellular energy levels, e.g., rich glucose conditions, SNF1 is not ADP bound and is exposed to dephophorylation resulting in low SNF1 activity (Mayer et al., 2011). Hence, a partial recovery in intracellular energy levels could explain the SUC2-expression, as it should be reflected by an increased inhibitory activity of the Snf1/Mig1 pathway (decreased inhibitory activity on Mig1 by the SNF1 complex), ultimately resulting in reduced *SUC2*-expression. Furthermore, a partial recovery in intracellular energy levels is expected upon long-term glucose starvation. For example, upon starvation reduced Snf1/Mig1 pathway inhibitory activity results in metabolism of alternative carbon sources and turns off energy producing processes (Hedbacker and Carlson, 2008).

To investigate if feedback mechanism via energy levels is mechanistically possible, we constructed a simple dynamic model (**Figure 3A** and Equation 1). The potential recovery in energy levels, which is a consequence of genes activated by glucose starvation and low Snf1/Mig1 pathway inhibitory activity, was modeled via the production of the feedback cascade. Furthermore, as a recovery in energy levels should result in increased Snf1/Mig1 pathway inhibitory activity, the feedback cascade was modeled to promote inhibitory activity. When fitted using a NLME approach, the model captures the single-cell behavior of the *SUC2* expression. More specifically, the model captures the observed individuality (**Figure S1A**) and also, by sampling from the estimated parameter distribution, is capable of simulating the observed population behavior (**Figure 3B**). An interesting model parameter is the feedback time-delay, τ_f, whose estimated distribution is separated from zero (**Figure S1C**). This suggests a delayed feedback with respect to the glucose downshift.

Overall, the feedback cascade model (**Figure 3A**) is able to explain the observed *SUC2* reduction upon long-term glucose starvation. As the feedback is modeled to behave as a partial recovery in intracellular energy levels, the feedback might act mainly via Snf1. However, due to the simplicity of the model it is not possible to deduce if this is a sufficient mechanism, e.g., the feedback might also act via Mig1 (**Figure 1**).

3.2.1. NLME Outperforms STS for Data Rich in Observations

Considering that parameter estimation for single-cell time-lapse data is challenging, we explored and compared the performance of standard two stage (STS) (Karlsson et al., 2015) and non-linear mixed-effect (NLME) (Almquist et al., 2015; Karlsson et al., 2015; Llamosi et al., 2016; Fröhlich et al., 2019; Marguet et al., 2019) approaches. NLME is considered superior to STS when the data is not rich (Karlsson et al., 2015). However, the experimental *SUC2* data obtained in this work can be viewed as sufficiently rich according to the criteria used by Karlsson et al. (2015). This is because the noise appears small, as there is a clear signal to noise ratio (**Figure 2B**) and furthermore, the data is not sparse (124 cells, with 97 non-zero observations each). Consequently, first fitting the simple feedback model to each cell, and from the fitted parameters infer the population parameters might yield as accurate parameter estimates, as the more advanced NLME-framework (Karlsson et al., 2015). It should be noted that when a system is perturbed by an external stimuli that yields a small effect, NLME outperforms the STS-approach (Karlsson et al., 2015). However, the observed *SUC2* data has no such stimuli, making this criteria irrelevant here.

Although both approaches appear to produce almost equally good individual fits (NLME slightly is better) to the observed *SUC2* data (**Figures S1A,D**), our analysis suggests that the STS-approach to a larger degree overestimates the variability when simulating the observed population behavior (**Figures 3B,C**). Furthermore, the STS fit does not capture the decrease in intensity for upper quantile equally well as the NLME approach. This suggests that the STS-approach estimates the parameter distribution incorrectly. Consequently, the STS-estimated distribution is unsuitable to use for further analysis, like model extrapolations. This is non-ideal from a computational perspective, as the computational times vary for two approaches (2.2 h to run the easily parallelizable STS estimation on a single core, compared to 5.0 h for the NLME estimation on eight cores). The difference when simulating the population behavior, might be due to outliers in the individual parameter estimates for the STS-approach. For example, the distribution assumption of k_6 is violated due to outliers in the STS-approach (**Figure S1E**).

Overall, the STS-approach is less suitable compared to NLME when simulating the observed population behavior, despite that both approaches have almost equally good individual fits.

3.3. SNF1 Is Central in Regulating *SUC2* Expression and Mig1 Nuclear Localization

The feedback cascade model (**Figure 3A**), does not reveal on which components of the Snf1/Mig1 pathway a *SUC2*-regulating feedback might act, but suggests a partial recovery in intracellular energy levels. Thus, a candidate is that the feedback acts mainly via the believed to be energy regulated (Rubenstein et al., 2008; Chandrashekarappa et al., 2013), SNF1-complex. A criteria for the feedback to act via SNF1, is that *SUC2* should be strongly regulated in a SNF1-dependent manner. To investigate this, we examined *SUC2* expression by measuring invertase activity under glucose rich and limited conditions (**Figure 4B**). Wild type (WT) cells showed an increased amount of glucose formed under glucose depletion compared to high glucose conditions. Absence of the *SNF1* gene resulted in decrease of the invertase activity regardless of the glucose presence. This supports previous findings that *SUC2* expression is regulated in a SNF1-dependent manner (Carlson et al., 1981; Neigeborn and Carlson, 1984). To be able to construct a model that examines the possibility of a *SUC2* regulating feedback acting via the SNF1-complex, we investigated the role of Snf1 in Mig1 localization. This was done by examining Mig1 localization when Snf1 kinase activity is inhibited (**Figure 4A**). We introduced *SNF1-I132G*, a PP1 analog-sensitive version of Snf1, into the cells with genomically integrated Mig1-GFP fusion as well as Nrd1-mCherry as a nuclear localization reporter. An *I132G*mutation at the ATP-binding pocket of Snf1 generates a novel structure sensitive to 1NM-PP1, an ATP competitive kinase inhibitor (Knight and Shokat, 2007; Rubenstein et al., 2008). Functionality of the analog-sensitive version of Snf1 has already been previously reported (Rubenstein et al., 2008; Shashkova et al., 2017). Upon glucose limitation, incubation of cells expressing *SNF1-I132G* with 25 μM 1NM-PP1 resulted in Mig1 retention in the nucleus, while the wild type Snf1 was irresponsive to the inhibitor. This is consistent with previous observations that Snf1 kinase activity is key for Mig1 nuclear export (Shashkova et al., 2017; Wollman et al., 2017).

Overall, experimental data suggests that *SUC2* is regulated in a SNF1-dependent manner. Furthermore, our data confirms that Snf1 activity is key for Mig1 nuclear export (Shashkova et al., 2017).

3.4. Modeling Reveals Potential Feedback Mechanism via Phosphorylated SNF1 After Initial Activation

Having the Mig1 localization and invertase activity data (**Figures 4A,B**), we were able to investigate the hypothesis that *SUC2* is mainly regulated in a SNF1-dependent manner upon long term glucose starvation by constructing a new dynamic model (**Figure 4D** and Equation 3). Considering that *SUC2* is regulated by SNF1 via Mig1 (**Figure 1**) we included the phosphorylated and dephosphorylated forms of nuclear Mig1. Further, it has been observed that upon a glucose downshift, a majority of Mig1 moves out of the nucleus (Treitel et al., 1998; Delyon et al., 1999), however, the mechanism behind this behavior is not known. To account for this observation, the transport behavior of Mig1 was included into the model by a sigmoid function (Equations 3b,c), which was parameterized to match observed Mig1 behavior (**Figure S3**). The potential partial recovery in energy levels, which is a consequence of genes activated by glucose starvation and high phosphorylated SNF1 activity (Hedbacker and Carlson, 2008), was modeled via the feedback mediating component whose production is promoted by high expression of Snf1/Mig1-controlled genes. Lastly, owing to available data (**Figure 4**), knowledge of the Snf1/Mig1 system (Shashkova et al., 2017), and that this model is more detailed regarding the Snf1/Mig1 pathway compared to the feedback cascade model (**Figure 3A**), the SNF1 related state variable was modeled to correspond to phosphorylated SNF1 (for detailed motivation see section 2.6). Thus, intracellular energy levels are reflected in the model by the activity of phosphorylated SNF1 (section 3.2). Hence, the feedback component which is promoted by a partial recovery in intracellular energy levels is modeled to inhibit the SNF1 related state variable (**Figure 4D**), in contrast to feedback cascade model where the feedback promotes inhibitory Snf1/Mig1 activity (**Figure 3A**).

Similarly to the first model, this model captures the observed individuality in *SUC2* data using a NLME approach as well as observed population behavior (**Figure 4E** and **Figure S2A**). It further captures that total nuclear Mig1 (phosohorylated + dephosphorylated) moves out of the nucleus upon glucose starvation (**Figure S2B**). Also, the model suggests that the total amount of nuclear Mig1 partially recovers upon long term glucose starvation (**Figure S2B**), which has been shown before (Dalal et al., 2014; Lin et al., 2015), due to decrease in nuclear export rate owing to the feedback acting on phosphorylated SNF1 (**Figure S2C**). Deleting the SNF1 component in the model by setting it to zero, leads to a constant low *SUC2* expression upon glucose starvation (**Figure 4F**). This matches the experimentally observed low *SUC2* activity when SNF1 is deleted (**Figure 4B**). However, the model suggests higher *SUC2* expression than observed in experimental data under the same conditions. To see further effects of deletions in the model, the ability of the model to mediate the feedback was removed by setting the feedback mediating component Y to zero (**Figure 4F**). This resulted in a strong increase of *SUC2*-expression compared to wild type. Deleting SNF1 and feedback mediating component yields the same result as when only deleting SNF1. This is expected, as the model assumes external glucose signals to be mediated to *SUC2* solely via the SNF1 complex.

Overall, the feedback mediated model (**Figure 4D**) is able to explain the experimental data reporting a reduction in *SUC2* expression upon long-term glucose starvation. As phosphorylated SNF1 is regulated via intracellular energy levels (section 3.2), this suggests that *SUC2* expression decreases due to a partial recovery in intracellular energy levels.

4. DISCUSSION

Nutrient sensing pathways are playing an important role in cellular response to different energy levels. Current understanding of this response only results in monotonic behavior of the *SUC2*-promoter upon starvation. However, our single-cell microfluidics data show that *SUC2* expression decreases in the long-term. To investigate the regulation of the *SUC2*-promoter upon long-term glucose starvation we have combined fluorescence microscopy and microfluidics data together with non-linear mixed effect modeling. The single-cell time-lapse data show that after an initial activation time a negative feedback takes place in the nutrient signaling network reducing the expression of the *SUC2* promoter (**Figure 2B**). Our invertase assay and microscopy data confirm that *SUC2* is regulated in a SNF1-dependent manner and that Snf1 activity is key for Mig1 nuclear export (**Figures 4A,B**). Finally, we propose via dynamic modeling that the decrease in SUC2 expression is due to a partial recovery in intracellular levels which results in a feedback that acts on the SNF1-complex (**Figures 4D,E**).

Without feedback, the *SUC2* would continue to increase, resulting in an excessive amount of invertase protein considering the energy supply and demand of the cell. This would result in the uneconomic use of cellular resources. Therefore, it is necessary to regulate the level of *SUC2* expression according to the cellular demand. Signals arising from changes in metabolic flux can be used to regulate invertase production (Litsios et al., 2018). Our feedback mediated model suggests that the expression of the target genes in the Snf1/Mig1 pathway is regulated through a feedback loop acting via a potential feedback mediating component (Y) (**Figure 4D**). We suggest that this component controls energy supply, produced through cellular metabolic flux. By coupling the metabolic flux with the production of enzymes, a stable and fast re-balancing of the enzymatic protein supply and demand can be created, which leads to optimal energy homeostasis. As it has been shown that the ratio of ADP/ATP in the cell controls the dephosphorylation of Snf1 in yeast (Mayer et al., 2011; Xiao et al., 2011; Chandrashekarappa et al., 2013), this ratio could be the sensor which couples metabolic flux (energy levels) in the cell with production of invertase, and thereby the mechanism responsible for the negative feedback. Also, other known mechanisms could be responsible for the negative feedback, such as the Reg1 phosphorylation by PKA, which activity is controlled by secondary messenger and ATP-derivative cAMP (Castermans et al., 2012). Further could the negative feedback also involve other proteins targeted by the Snf1 pathway, such as Cat8, Adr1, and Sip4. These proteins are directly phosphorylated by Snf1 kinase and the expression of the encoding genes is controlled by Mig1 (DeVit et al., 1997). Cat8, Adr1, and Sip4 are involved in the cellular reprogramming during the diauxic shift and through this role influence the cell energy-metabolism (Vincent and Carlson, 1998; Haurie et al., 2001; Young et al., 2003). The alteration caused in the metabolism by these protein could be pivotal in the changing behavior of the *SUC2* expression.

The Glc7-Reg1 phosphatase is required for Snf1 dephosphorylation (Rubenstein et al., 2008). At the same time, Snf1 itself acts on Reg1 and prevents its association with the Glc7 subunit for the formation of the functional phosphatase (Sanz et al., 2000). This loop makes Reg1 a potential candidate to be involved in the energy regulated feedback proposed by our modeling (**Figure 4D**). Our simulations on cells deficient in potential feedback mediated component (Y) suggest an increase in the *SUC2* expression compared to the wild type (**Figure 4F**). We tested how the target genes are affected in yeast cells carrying the *Reg1* deletion. Our experimental data suggests an increase in invertase activity upon *REG1* deletion compared to the WT (**Figure 4C**). Furthermore, invertase activity on the cells without both Reg1 and Snf1 shows reduced invertase-activity, which is in agreement with the reduced *SUC2* expression when deleting *SNF1* and feedback mediated component (Y) in the model. This suggests that Reg1 is a central part of the feedback. However, as the feedback mediation in the model encompasses all potential energy regulated components that affects SNF1 activity, more components than Reg1 are likely involved in the feedback.

Our feedback meditating model suggests that the fine-tuning of expression, after the initial strong activation, of the target genes in the Snf1/Mig1 pathway is regulated through a feedback loop acting via a potential feedback mediating component (Y). Mig1 regulates genes essential for utilization of carbon sources (Lutfiyya et al., 1998), hence, Mig1 participates in controlling energy metabolism in the cell. This is another evidence supporting Reg1 being central part of the feedback mediated component as the intracellular energy levels have been shown to play an important role in the activation of Snf1, thus, its communication with the Glc7-Reg1 phosphatase (Rubenstein et al., 2008).

Deletion of Snf1 in the feedback mediated model resulted in a stable *SUC2* expression at a level similar to simulated wild type at time 0, which corresponds closely to *SUC2* expression at 4% glucose (**Figure 4F**). However, the invertase activity assay shows that the activity in the *SNF1* deletion strain is lower then the WT at 4% glucose (**Figure 4B**). This highlights that the model does not fully capture the behavior of the *SUC2* expression at 4% glucose. A probable cause of this discrepancy is the simplicity of the model. Other pathways have been shown to influence the *SUC2* expression and cross-talk has been suggested to be ubiquitous in the nutrient signaling system (Kayikci and Nielsen, 2015; Shashkova et al., 2017). A larger model, opposed to the small-scale model in this work, could include other pathways known to be able to influence the expression of *SUC2*. Larger mathematical models have been made (Kayikci and Nielsen, 2015; Welkenhuysen et al., 2018), however they are Boolean models and consequently cannot capture the dynamic, single-cell behavior of cells exposed to several environmental conditions. Due to this inherit drawback of Boolean models and the connectivity of the nutrient sensing pathways, a large-scale single-cell model is probably needed to fully understand the Snf1/Mig1 pathway dynamics.

Parameter estimation would pose a considerable challenge in constructing a large scale, mechanistic single-cell model of the nutrient sensing network. Here, we compared two estimation methods for single-cell time-lapse data, STS and NLME. Although our data is rich in observations, NLME outperformed STS when estimating the population parameter distribution (**Figure 3**). As discussed by Almquist et al. (2015), this is probably due to some cells not carrying sufficient information to properly estimate all parameters. Hence, parameters like k_6 can take extreme values for certain cells (**Figure S1E**), ultimately resulting in bad estimates of the population parameters. Our result thus highlight that data rich in observations, is not equal to data that is optimally sampled for each cell. Consequently, it is far from guaranteed that STS, although it happens (Karlsson et al., 2015), equals NLME in performance for observation rich data. This fact, combined with previous underperformance (Almquist et al., 2015; Llamosi et al., 2016), suggests that the STS-approach is not a preferred method. However, this does not mean that the current NLME framework should be the preferred method for a large-scale model built on time-lapse data. For example, here we show that NMLE is computationally demanding even for a small model. The global two-stage (GTS) approach has been proposed as an alternative to NLME (Dharmarajan et al., 2019). However, GTS currently cannot handle multi-experiment data (Loos and Hasenauer, 2019), and it is questionable if a large model can be calibrated using single-experiment data. Overall, these shortcomings highlight that further development in parameter estimation methods is required for constructing large-scale mechanistic single-cell models.

In summary, our systems biology approach suggests that *SUC2* expression decrease upon long-term glucose starvation is due to a partial recovery in intracellular energy levels acting on the SNF1-complex.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in FigShare repository. The *SUC2* expression data for this study: https://figshare.com/s/d846d38177821c3a2c4e. The time-lapse microscopy images for this study: https://figshare.com/s/6544469f4cf99fc0d862.

AUTHOR CONTRIBUTIONS

SP developed the mathematical models, performed the simulations, and mathematical analysis. NW and SS planned and performed the experimental part of the work. NW and MC conceived the research. SP, NW, SS, and MC wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Swedish Foundation for Strategic Research (Grant No. FFL15-0238) to MC and FEMS Research and Travel (Grant No. FEMS-GO-2018-213) to SS.

ACKNOWLEDGMENTS

We would like to thank all past and present members of the CvijovicLab for valuable input and careful reading of the manuscript and Hohmann lab for providing strains and plasmids.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00954/full#supplementary-material

REFERENCES

Ahuatzi, D., Riera, A., Peláez, R., Herrero, P., and Moreno, F. (2007). Hxk2 regulates the phosphorylation state of Mig1 and therefore its nucleocytoplasmic distribution. *J. Biol. Chem.* 282, 4485–4493. doi: 10.1074/jbc.M606854200

Almquist, J., Bendrioua, L., Adiels, C. B., Goksör, M., Hohmann, S., and Jirstrand, M. (2015). A nonlinear mixed effects approach for modeling the cell-to-cell variability of Mig1 dynamics in yeast. *PLoS ONE* 10:124050. doi: 10.1371/journal.pone.0124050

Atkins, P. W. P. W., De Paula, J., and Friedman, R. (2013). *Physical Chemistry: Quanta, Matter, and Change, 2 rev Edn*. Oxford: Oxford University Press.

Bendrioua, L., Smedh, M., Almquist, J., Cvijovic, M., Jirstrand, M., Goksör, M., et al. (2014). Yeast AMP-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *J. Biol. Chem.* 289, 12863–12875. doi: 10.1074/jbc.M114.547976

Berben, G., Dumont, J., Gilliquet, V., Bolle, P.-A., and Hilger, F. (1991). The YDp plasmids: a uniform set of vectors bearing versatile gene disruption cassettes forSaccharomyces cerevisiae. *Yeast* 7, 475–477. doi: 10.1002/yea.320070506

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: a fresh approach to numerical computing. *SIAM* 59, 65–98. doi: 10.1137/141000671

Broach, J. R. (2012). Nutritional control of growth and development in yeast. *Genetics* 192, 73–105. doi: 10.1534/genetics.111.135731

Carlson, M. (1999). Glucose repression in yeast. *Curr. Opin. Microbiol.* 2, 202–207. doi: 10.1016/S1369-5274(99)80035-6

Carlson, M., and Botstein, D. (1982). Two differentially regulated mRNAs with different 5′ ends encode secreted with intracellular forms of yeast invertase. *Cell* 28, 145–154. doi: 10.1016/0092-8674(82)90384-1

Carlson, M., Osmond, B. C., and Botstein, D. (1981). Mutants of yeast defective in sucrose utilization. *Genetics* 98, 25–40.

Castermans, D., Somers, I., Kriel, J., Louwet, W., Wera, S., Versele, M., et al. (2012). Glucose-induced posttranslational activation of protein phosphatases PP2A and PP1 in yeast. *Cell Res.* 22, 1058–1077. doi: 10.1038/cr.2012.20

Celenza, J. L., and Carlson, M. (1986). A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* 233, 1175–1180. doi: 10.1126/science.3526554

Chan, P. L., Jacqmin, P., Lavielle, M., McFadyen, L., and Weatherley, B. (2011). The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. *J. Pharmacokinet. Pharmacodyn.* 38, 41–61. doi: 10.1007/s10928-010-9175-z

Chandrashekarappa, D. G., McCartney, R. R., and Schmidt, M. C. (2013). Ligand binding to the AMP-activated protein kinase active site mediates protection of the activation loop from dephosphorylation. *J. Biol. Chem.* 288, 89–98. doi: 10.1074/jbc.M112.422659

Cohen-Saidon, C., Cohen, A. A., Sigal, A., Liron, Y., and Alon, U. (2009). Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol. Cell* 36, 885–893. doi: 10.1016/j.molcel.2009.11.025

Dalal, C. K., Cai, L., Lin, Y., Rahbar, K., and Elowitz, M. B. (2014). Pulsatile dynamics in the yeast proteome. *Curr. Biol.* 24, 2189–2194. doi: 10.1016/j.cub.2014.07.076

Davidian, M., and Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: an overview and update. *J. Agric. Biol. Environ. Stat.* 8, 387–419. doi: 10.1198/1085711032697

Delyon, B., Lavielle, M., of Statistics, E. M. A., and 1999, U. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Stat.* 27, 94–128. doi: 10.1214/aos/1018031103

DeVit, M. J., Waddle, J. A., and Johnston, M. (1997). Regulated nuclear translocation of the Mig1 glucose repressor. *Mol. Biol. Cell* 8, 1603–1618. doi: 10.1091/mbc.8.8.1603

Dharmarajan, L., Kaltenbach, H. M., Rudolf, F., and Stelling, J. (2019). A simple and flexible computational framework for inferring sources of heterogeneity from single-cell dynamics. *Cell Syst.* 8, 15–26.e11. doi: 10.1016/j.cels.2018.12.007

Elbing, K., McCartney, R. R., and Schmidt, M. C. (2006). Purification and characterization of the three Snf1-activating kinases of Saccharomyces cerevisiae. *Biochem. J.* 393, 797–805. doi: 10.1042/BJ20051213

Fröhlich, F., Reiser, A., Fink, L., Woschée, D., Ligon, T., Theis, F. J., et al. (2019). Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. *NPJ Syst. Biol. Appl.* 5, 1–12. doi: 10.1038/s41540-018-0079-7

Ghillebert, R., Swinnen, E., Wen, J., Vandesteene, L., Ramon, M., Norga, K., et al. (2011). The AMPK/SNF1/SnRK1 fuel gauge and energy regulator: structure, function and regulation. *FEBS J.* 278, 3978–3990. doi: 10.1111/j.1742-4658.2011.08315.x

Gietz, R. D., and Schiestl, R. H. (2007). Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 1–4. doi: 10.1038/nprot.2007.17

Gordon, A., Colman-Lerner, A., Chin, T. E., Benjamin, K. R., Yu, R. C., and Brent, R. (2007). Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Methods* 4, 175–181. doi: 10.1038/nmeth1008

Haurie, V., Perrot, M., Mini, T., Jeno, P., Sagliocco, F., and Boucherie, H. (2001). The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in Saccharomyces cerevisiae. *J. Biol. Chem.* 276, 76–85. doi: 10.1074/jbc.M008752200

Hedbacker, K., and Carlson, M. (2008). SNF1/AMPK pathways in yeast. *Front. Biosci.* 13:2408. doi: 10.2741/2854

Hong, S. P., Leiper, F. C., Woods, A., Carling, D., and Carlson, M. (2003). Activation of yeast Snf1 and mammalian AMP-activated protein kinase by upstream kinases. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8839–8843. doi: 10.1073/pnas.1533136100

Jiang, R., and Carlson, M. (1996). Glucose regulates protein interactions within the yeast SNF1 protein kinase complex. *Genes Dev.* 10, 3105–3115. doi: 10.1101/gad.10.24.3105

Johnson, S. (2020). *The NLopt Nonlinear-Optimization Package*. Available online at: http://github.com/stevengj/nlopt

Karlsson, M., Janzén, D. L., Durrieu, L., Colman-Lerner, A., Kjellsson, M. C., and Cedersund, G. (2015). Nonlinear mixed-effects modelling for single cell estimation: when, why, and how to use it. *BMC Syst. Biol.* 9:52. doi: 10.1186/s12918-015-0203-x

Kayikci, Ö., and Nielsen, J. (2015). Glucose repression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15, 1–8. doi: 10.1093/femsyr/fov068

Knight, Z. A., and Shokat, K. M. (2007). Chemical genetics: where genetics and pharmacology meet. *Cell* 128, 425–430. doi: 10.1016/j.cell.2007.01.021

Kuhn, E., and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Stat. Data Anal.* 49, 1020–1038. doi: 10.1016/j.csda.2004.07.002

Lavielle, M., and Ribba, B. (2016). Enhanced method for diagnosing pharmacometric models: random sampling from conditional distributions. *Pharma. Res.* 33, 2979–2988. doi: 10.1007/s11095-016-2020-3

Limpert, E., Stahel, W. A., and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience* 51, 341–352. doi: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Lin, Y., Sohn, C. H., Dalal, C. K., Cai, L., and Elowitz, M. B. (2015). Combinatorial gene regulation by modulation of relative pulse timing. *Nature* 527, 54–58. doi: 10.1038/nature15710

Litsios, A., Ortega, Á, D., Wit, E. C., and Heinemann, M. (2018). Metabolic-flux dependent regulation of microbial physiology. *Curr. Opin. Microbiol.* 42, 71–78. doi: 10.1016/j.mib.2017.10.029

Lixoft (2019). *Monolix Version 2019R2*. Antony: Lixoft SAS. Available online at: http://lixoft.com/products/monolix/

Llamosi, A., Gonzalez-Vargas, A. M., Versari, C., Cinquemani, E., Ferrari-Trecate, G., Hersen, P., et al. (2016). What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. *PLoS Comput. Biol.* 12:e1004706. doi: 10.1371/journal.pcbi.1004706

Loos, C., and Hasenauer, J. (2019). Mathematical modeling of variability in intracellular signaling. *Curr. Opin. Syst. Biol.* 16, 17–24. doi: 10.1016/j.coisb.2019.10.020

Lutfiyya, L. L., Iyer, V. R., DeRisi, J., DeVit, M. J., Brown, P. O., and Johnston, M. (1998). Characterization of three related glucose repressors and genes they regulate in Saccharomyces cerevisiae. *Genetics* 150, 1377–1391.

Marguet, A., Lavielle, M., and Cinquemani, E. (2019). Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data. *Bioinformatics* 35, i586–i595. doi: 10.1093/bioinformatics/btz378

Mayer, F. V., Heath, R., Underwood, E., Sanders, M. J., Carmena, D., McCartney, R. R., et al. (2011). ADP regulates SNF1, the saccharomyces cerevisiae homolog of AMP-activated protein kinase. *Cell Metab.* 14, 707–714. doi: 10.1016/j.cmet.2011.09.009

McCartney, R. R., Garnar-Wortzel, L., Chandrashekarappa, D. G., and Schmidt, M. C. (2016). Activation and inhibition of Snf1 kinase activity by phosphorylation within the activation loop. *Biochim. Biophys. Acta* 1864, 1518–1528. doi: 10.1016/j.bbapap.2016.08.007

McCartney, R. R., and Schmidt, M. C. (2001). Regulation of Snf1 kinase. Activation requires phosphorylation of threonine 210 by an upstream kinase as well as a distinct step mediated by the Snf4 subunit. *J. Biol. Chem.* 276, 36460–36466. doi: 10.1074/jbc.M104418200

Nath, N., McCartney, R. R., and Schmidt, M. C. (2003). Yeast Pak1 kinase associates with and activates Snf1. *Mol. Cell. Biol.* 23, 3909–3917. doi: 10.1128/MCB.23.11.3909-3917.2003

Neigeborn, L., and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* 108, 845–858.

Noble, W. S. (2009). A quick guide to organizing computational biology projects. *PLoS Comput. Biol.* 5:e1000424. doi: 10.1371/journal.pcbi.1000424

Powell, M. J. (2009). *The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives*. Technical report, Department of Applied Mathematics and Theoretical Physics, Cambridge, United Kingdom.

Rackauckas, C., and Nie, Q. (2017). DifferentialEquations.jl–a performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Soft.* 5:15. doi: 10.5334/jors.151

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., et al. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25, 1923–1929. doi: 10.1093/bioinformatics/btp358

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE* 8:e74335. doi: 10.1371/annotation/ea0193d8-1f7f-492a-b0b7-d877629fdcee

Rubenstein, E. M., McCartney, R. R., Zhang, C., Shokat, K. M., Shirra, M. K., Arndt, K. M., et al. (2008). Access denied: Snf1 activation loop phosphorylation is controlled by availability of the phosphorylated threonine 210 to the PP1 phosphatase. *J. Biol. Chem.* 283, 222–230. doi: 10.1074/jbc.M707957200

Ruiz, A., Liu, Y., Xu, X., and Carlson, M. (2012). Heterotrimer-independent regulation of activation-loop phosphorylation of Snf1 protein kinase involves two protein phosphatases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8652–8657. doi: 10.1073/pnas.1206280109

Ruiz, A., Xu, X., and Carlson, M. (2011). Roles of two protein phosphatases, Reg1-Glc7 and Sit4, and glycogen synthesis in regulation of SNF1 protein kinase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6349–6354. doi: 10.1073/pnas.1102758108

Ruiz, A., Xu, X., and Carlson, M. (2013). Ptc1 protein phosphatase 2C contributes to glucose regulation of SNF1/AMP-activated protein kinase (AMPK) in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 288, 31052–31058. doi: 10.1074/jbc.M113.503763

Sanz, P., Alms, G. R., Haystead, T. A. J., and Carlson, M. (2000). Regulatory interactions between the Reg1-Glc7 protein phosphatase and the Snf1 protein kinase. *Mol. Cell. Biol.* 20, 1321–1328. doi: 10.1128/MCB.20.4.1321-1328.2000

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019

Schmidt, M. C., and McCartney, R. R. (2000). Beta-subunits of Snf1 kinase are required for kinase function and substrate definition. *EMBO J.* 19, 4936–4943. doi: 10.1093/emboj/19.18.4936

Shashkova, S., and Leake, M. C. (2017). Single-molecule fluorescence microscopy review: shedding new light on old problems. *Biosci. Rep.* 37:BSR20170031. doi: 10.1042/BSR20170031

Shashkova, S., Welkenhuysen, N., and Hohmann, S. (2015). Molecular communication: crosstalk between the Snf1 and other signaling pathways. *FEMS Yeast Res.* 15, 1–10. doi: 10.1093/femsyr/fov026

Shashkova, S., Wollman, A. J., Leake, M. C., and Hohmann, S. (2017). The yeast Mig1 transcriptional repressor is dephosphorylated by glucose-dependent and -independent mechanisms. *FEMS Microbiol. Lett.* 364, 1–9. doi: 10.1093/femsle/fnx133

Smith, F. C., Davies, S. P., Wilson, W. A., Carling, D., and Hardie, D. (1999). The SNF1 kinase complex from *Saccharomyces cerevisiae* phosphorylates the transcriptional repressor protein Mig1p *in vitro* at four sites within or near regulatory domain 1. *FEBS Lett.* 453, 219–223. doi: 10.1016/S0014-5793(99)00725-5

Treitel, M. A., and Carlson, M. (1995). Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3132–3136. doi: 10.1073/pnas.92.8.3132

Treitel, M. A., Kuchin, S., and Carlson, M. (1998). Snf1 protein kinase regulates phosphorylation of the Mig1 repressor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 18, 6273–6280. doi: 10.1128/MCB.18.11.6273

Vincent, O., and Carlson, M. (1998). Sip4, a Snf1 kinase-dependent transcriptional activator, binds to the carbon source-responsive element of gluconeogenic genes. *EMBO J.* 17, 7002–7008. doi: 10.1093/emboj/17.23.7002

Welkenhuysen, N., Adiels, C. B., Goksör, M., and Hohmann, S. (2018). Applying microfluidic systems to study effects of glucose at single-cell level. *Methods Mol. Biol.* 1713, 109–121. doi: 10.1007/978-1-4939-7507-5_9

Welkenhuysen, N., Borgqvist, J., Backman, M., Bendrioua, L., Goksör, M., Adiels, C. B., et al. (2017). Single-cell study links metabolism with nutrient signaling and reveals sources of variability. *BMC Syst. Biol.* 11:59. doi: 10.1186/s12918-017-0435-z

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., and Teal, T. K. (2017). Good enough practices in scientific computing. *PLoS Computat. Biol.* 13:e1005510. doi: 10.1371/journal.pcbi.1005510

Wollman, A. J., Shashkova, S., Hedlund, E. G., Friemann, R., Hohmann, S., and Leake, M. C. (2017). Transcription factor clusters regulate genes in eukaryotic cells. *eLife* 6:e27451. doi: 10.7554/eLife.27451

Wu, J., and Trumbly, R. J. (1998). Multiple regulatory proteins mediate repression and activation by interaction with the yeast Mig1 binding site. *Yeast* 14, 985–1000.

Xiao, B., Sanders, M. J., Underwood, E., Heath, R., Mayer, F. V., Carmena, D., et al. (2011). Structure of mammalian AMPK and its regulation by ADP. *Nature* 472, 230–233. doi: 10.1038/nature09932

Young, E. T., Dombek, K. M., Tachibana, C., and Ideker, T. (2003). Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *J. Biol. Chem.* 278, 26146–26158. doi: 10.1074/jbc.M301981200

Zhang, Y., McCartney, R. R., Chandrashekarappa, D. G., Mangat, S., and Schmidt, M. C. (2011). Reg1 protein regulates phosphorylation of all three Snf1 isoforms but preferentially associates with the Gal83 isoform. *Eukaryot. Cell* 10, 1628–1636. doi: 10.1128/EC.05176-11

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 30 September 2020
doi: 10.3389/fphys.2020.558606

Computational Verification of Large Logical Models—Application to the Prediction of T Cell Response to Checkpoint Inhibitors

Céline Hernandez [1†], Morgane Thomas-Chollier [1,2], Aurélien Naldi [1†] and Denis Thieffry [1*]*

[1] Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France, [2] Institut Universitaire de France, Paris, France

Edited by:
Jianhua Xing,
University of Pittsburgh, United States

Reviewed by:
Carlos F. Lopez,
Vanderbilt University, United States
Andrei Zinovyev,
Institut Curie, France

***Correspondence:**
Aurélien Naldi
aurelien.naldi@gmail.com
Denis Thieffry
denis.thieffry@bio.ens.psl.eu

†Present address:
Céline Hernandez,
Université Paris-Saclay, CEA, CNRS,
Institute for Integrative Biology of the
Cell (I2BC), Gif-sur-Yvette, France
Aurélien Naldi,
EP Lifeware, INRIA
Saclay–Ile-de-France, Palaiseau,
France

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: *03 May 2020*
Accepted: *19 August 2020*
Published: *30 September 2020*

Citation:
Hernandez C, Thomas-Chollier M,
Naldi A and Thieffry D (2020)
Computational Verification of Large
Logical Models—Application to the
Prediction of T Cell Response to
Checkpoint Inhibitors.
Front. Physiol. 11:558606.
doi: 10.3389/fphys.2020.558606

At the crossroad between biology and mathematical modeling, computational systems biology can contribute to a mechanistic understanding of high-level biological phenomenon. But as knowledge accumulates, the size and complexity of mathematical models increase, calling for the development of efficient dynamical analysis methods. Here, we propose the use of two approaches for the development and analysis of complex cellular network models. A first approach, called "model verification" and inspired by unitary testing in software development, enables the formalization and automated verification of validation criteria for whole models or selected sub-parts. When combined with efficient analysis methods, this approach is suitable for continuous testing, thereby greatly facilitating model development. A second approach, called "value propagation," enables efficient analytical computation of the impact of specific environmental or genetic conditions on the dynamical behavior of some models. We apply these two approaches to the delineation and the analysis of a comprehensive model for T cell activation, taking into account CTLA4 and PD-1 checkpoint inhibitory pathways. While model verification greatly eases the delineation of logical rules complying with a set of dynamical specifications, propagation provides interesting insights into the different potential of CTLA4 and PD-1 immunotherapies. Both methods are implemented and made available in the all-inclusive CoLoMoTo Docker image, while the different steps of the model analysis are fully reported in two companion interactive jupyter notebooks, thereby ensuring the reproduction of our results.

Keywords: T cell, checkpoint inhibitors, Boolean models, model verification, value propagation

INTRODUCTION

Recent technical developments have allowed scientists to study immunology and health-related issues from a variety of angles. For many diseases, especially for cancer, the current trend consists in aggregating data coming from different sources to gain a global view of cell, tissue, or organ dysfunction. Over the last decades, diverse mathematical frameworks have been proposed to seize a multiplicity of biological questions (Le Novère, 2015), including in immunology (Kaufman et al., 1985, 1999; Eftimie et al., 2016; Chakraborty, 2017). However, the increasing complexity of biological questions implies the development of more sophisticated models, which in turn bring serious computational challenges.

Among the mathematical approaches proposed for the modeling of cellular networks, the logical modeling framework is increasingly used. In particular, it has been successfully applied to immunology and cancer, leading to the creation of models encompassing dozens of components, some including many inputs components (Grieco et al., 2013; Abou-Jaoudé et al., 2014; Flobak et al., 2015; Oyeyemi et al., 2015). However, the large size of recent models hinders the complete exploration of their dynamical behavior through simulation, especially in non-deterministic settings.

To address these difficulties, we define and apply a *model verification* approach to systematically verify whether a model complies with a list of known properties. These properties are defined as model *specifications*, either at a local (i.e., for sub-models) or at a global level. This automated verification procedure fosters confidence during the development of a complex dynamical model and paves the way to the development of models with hundreds of nodes.

We further outline and apply a *value propagation* method, which enables the assessment of the impact of environmental or genetic constraints on the dynamical behavior of complex cellular networks.

These two complementary approaches can be applied to the development and analysis of large dynamical models, as illustrated in **Figure 1**. Noteworthy, they have been implemented in a multi-platform Docker image combining various complementary logical modeling and analysis tools (Naldi et al., 2018b). We further illustrate the power of these methods through the analysis of an original model. The different steps of analysis are fully reported in two companion interactive jupyter notebooks, available with the model on the GINsim website (http://ginsim.org/model/tcell-checkpoint-inhibitors-tcla4-pd1), thereby ensuring their reproducibility.

MODEL VERIFICATION

A Software Engineering Framework for Logical Model Building

One of the main features determining the interest of a model is its ability to accurately recapitulate salient biological knowledge. More precisely, this knowledge can be used in two complementary ways during the model building process. On the one hand, it is used to define the model architecture, specifying which biological entities need to be included and which interactions between these entities need to be encoded. On the other hand, biological knowledge entails dynamical properties that must be achieved by the resulting model, whether transitory or asymptotic, to account for biological observations. These properties induce satisfaction criteria and must be clearly specified for rigorous model assessment or comparison with other models. Failures to reproduce such properties need to be carefully documented, thereby providing a basis for further model improvement.

In the domain of logical modeling applied to cellular networks, various formal methods have already been proposed to verify dynamical properties. For example *stable states* (or *fixed points*, characterized by all components being steady at the same time) tentatively correspond to asymptotic properties that can used to assess the reproduction of known persistent biological behavior. More complex asymptotic behaviors include *cyclic attractors*, which can be approximate by the computation of so called *trap spaces*. Also called *stable motifs*, trap spaces are hypercubes in the state space such that all successors of all states in the hypercube also belong to it (for synchronous and asynchronous updatings, or any other updating). These hypercubes then provide an approximation of complex attractors. Trap spaces and stable states can be defined as results of a constraint solving system, enabling their efficient computation (Klarner et al., 2018). Their reachability however must be assessed separately, often using model checking or stochastic simulations, which requires longer computations.

Model checking techniques have been successfully applied to specify and verify temporal constraints on a model behavior (Monteiro and Chaouiya, 2012; Miskov-Zivanov et al., 2016; Traynard et al., 2016; Wang et al., 2016).

In any case, whatever the formalism chosen, the building of a complex dynamical model is intrinsically iterative, as its establishment is usually incremental and requires continuous testing and adjustment with reference to a growing body of biological knowledge.

In the field of software engineering, the similar need to repeatedly assess criteria of success or failure of a software program led to the development of powerful *software verification* techniques, and in particular to software testing (Myers, 1979), which main goal is to assess whether a software meets a series of well-defined requirements. More importantly, such assessments must be repeated as soon as a new piece of code or specification is added. Software testing aims to check whether newly introduced modification might break any of the previous performances. In particular, software verification includes the notion of *unit testing*, where suites of tests describe the expected behavior associated with individual units composing a program. This idea can be transposed from computer science to model building and has been successfully applied in the context of other modeling frameworks (Hoops et al., 2006; Lopez et al., 2013; Sarma et al., 2016; Boutillier et al., 2018), but not yet to logical modeling.

Here, we transpose the *unit testing* approach to integrate a comprehensive series of verifiable criteria, from the early stages of model conception, in order to automate the dynamical evaluation of logical models. The core idea is to split the biological knowledge on which a model is based into individual verifiable criteria that can be formalized as specifications (**Figures 1**, **2**). In this respect, individual units of knowledge, derived from the scientific literature or biological experiments, must be formulated into stable or dynamical properties. Each specification, coupling a property with an expected value, can serve as a basis to define a test case for a model. Testing such a specification amounts to compute an "observed" value based on the model and compare this value to the expected one.

In practice, the CoLoMoTo notebook environment (Naldi et al., 2018b) provides a Python API for several software

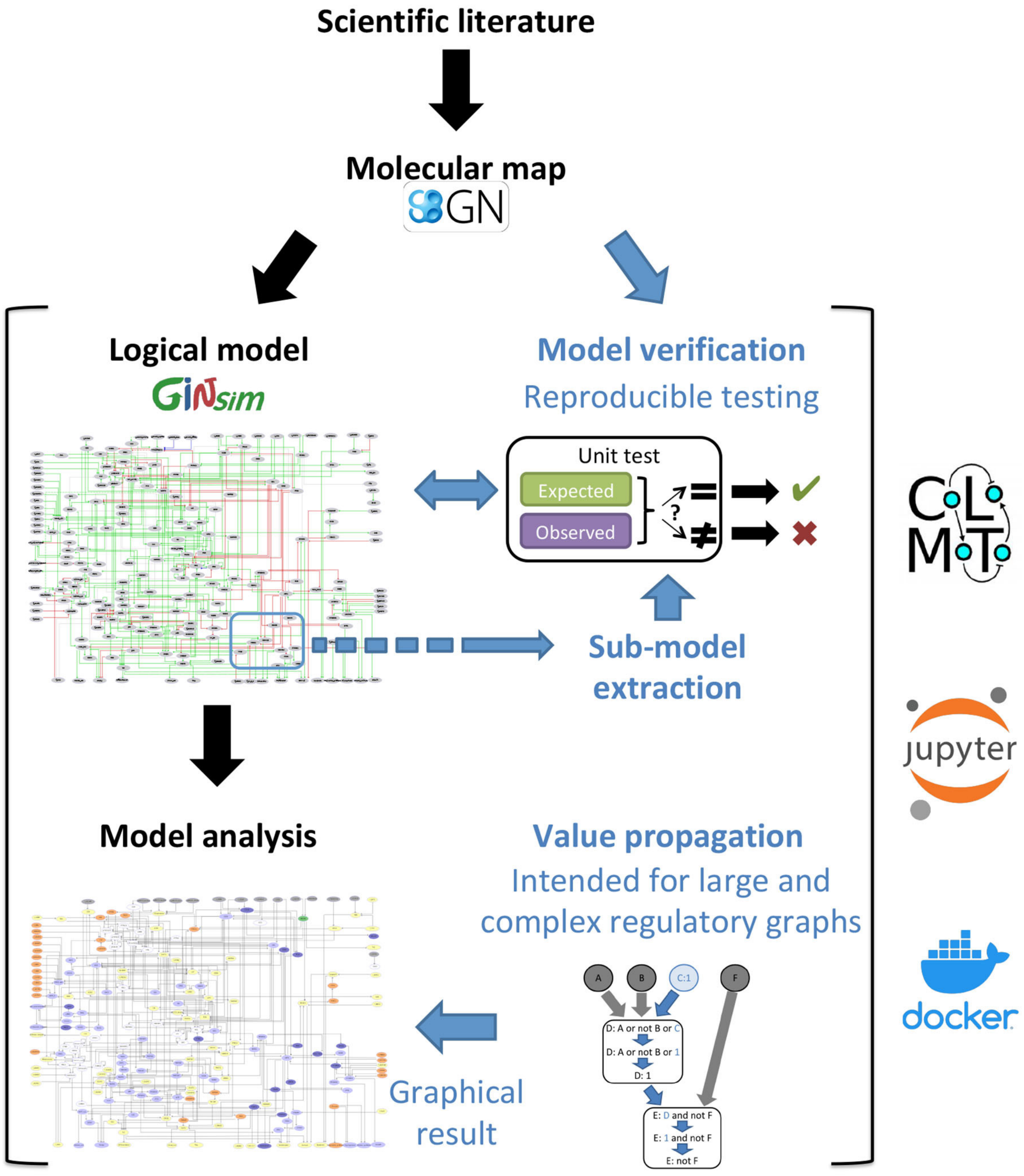
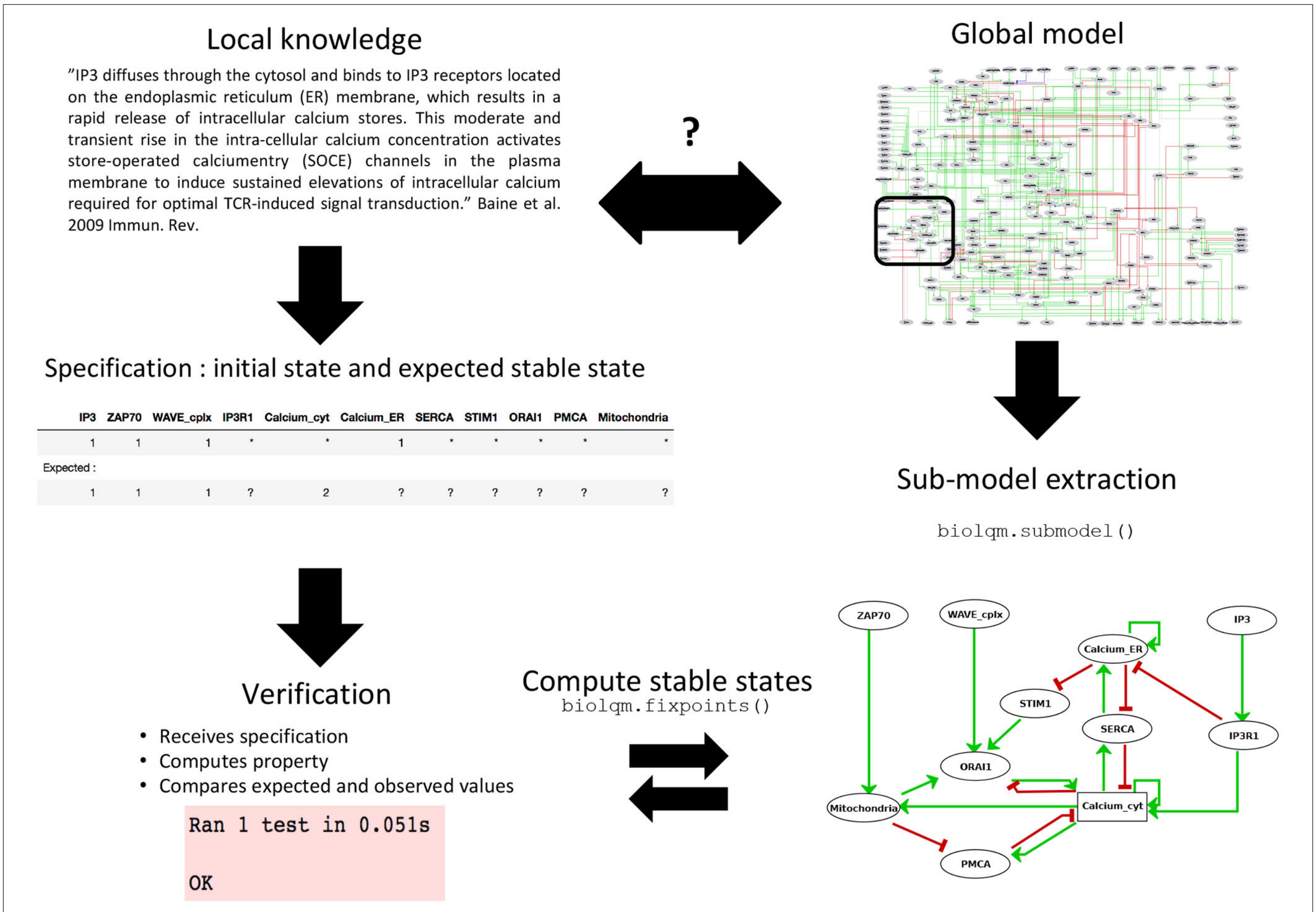


FIGURE 1 | Description of the proposed workflow for the development and analysis of dynamical logical models. The novel methods described in this article are emphasized with blue fonts. Starting with the delineation of a molecular map integrating the available scientific knowledge, we derive a regulatory graph and logical rules to generate a logical model, and induce dynamical specifications serving as test cases to verify the model. Moreover, when the available knowledge is specific to a smaller part of the regulatory graph, a sub-model us extracted to perform local tests. We further implemented an analysis and visualization method, called Value propagation, to assess the impact of environmental and genetic perturbations. **Figure 3** zooms into this part of the workflow and describes it in more details. The use of model verification, sub-model extraction and value propagation is illustrated in two reproducible and editable Jupyter notebooks, taking advantage of the CoLoMoTo Interactive notebook framework (Naldi et al., 2018b). This framework is available inside the CoLoMoTo Docker image together with packaged libraries for the analysis of dynamical logical models of biological networks.

tools, enabling the definition of a wide range of dynamical analyses for the computation of observed values. Individual test cases can be assembled into a library, also called *testing suite*. Existing tools and packages enabling software testing can then be applied to automatically assess whether a model satisfies (or not) a series of specifications. In this study,



"IP3 diffuses through the cytosol and binds to IP3 receptors located on the endoplasmic reticulum (ER) membrane, which results in a rapid release of intracellular calcium stores. This moderate and transient rise in the intra-cellular calcium concentration activates store-operated calciumentry (SOCE) channels in the plasma membrane to induce sustained elevations of intracellular calcium required for optimal TCR-induced signal transduction." Baine et al. 2009 Immun. Rev.

	IP3	ZAP70	WAVE_cplx	IP3R1	Calcium_cyt	Calcium_ER	SERCA	STIM1	ORAI1	PMCA	Mitochondria
	1	1	1	*	*	1	*	*	*	*	*
Expected :											
	1	1	1	?	2	?	?	?	?	?	?

- Receives specification
- Computes property
- Compares expected and observed values

```
Ran 1 test in 0.051s

OK
```

FIGURE 2 | Sub-model extraction for local model verification. When the available knowledge is fragmentary and covers the behavior of only a subset of components, verification becomes difficult at the global scale of the model. Based on this partial information, a series of specifications can still be defined for a sub-model which, after extraction using bioLQM's submodel() function, can then be rigorously tested.

we used the python package "unittest," taking advantage of its seamless integration into the CoLoMoTo interactive notebook. This unit-testing package is integrated by default into the recent versions of the Python standard library (http://python.org).

Local Verification of Sub-Models Can Cope With Sparsity of Biological Knowledge

Biological knowledge reported in the scientific literature is often insufficient to evaluate a comprehensive model, which may encompass hundreds of nodes. In particular, observations regarding component activity often relate to only a limited subset of nodes of the model. This greatly complicates the definition of specifications for the whole model.

Given a comprehensive model and a set of components of interest, one can extract a sub-model containing these core components, along with their associated logical rules. Components appearing in these logical rules but not in the selected set are considered as external inputs of the sub-model (**Figure 2**). This functionality has been implemented in the "submodel" function of the Java bioLQM library (Naldi, 2018) according to the following procedure.

Let $M = (V, f)$ be a model, where V is the set of components, and f the update function. For each c in V, f_c is the logical function associated to the component c and $R(c)$ is the set of its regulators (i.e., components that intervene in the logical rule). Given a list of selected components $C \subset V$:

1. $S = \emptyset$
2. for each component $c \in C$: $S = S + \{c\} + R(c)$
3. create the sub-model $M' = (S, f')$ such that for each component c in S: $f'_c = \begin{cases} f_c & \text{if } R(c) \subset S \\ c & \text{otherwise} \end{cases}$

As shown in the application below, the delineation of such sub-models can greatly facilitate the definition and verification of local specifications.

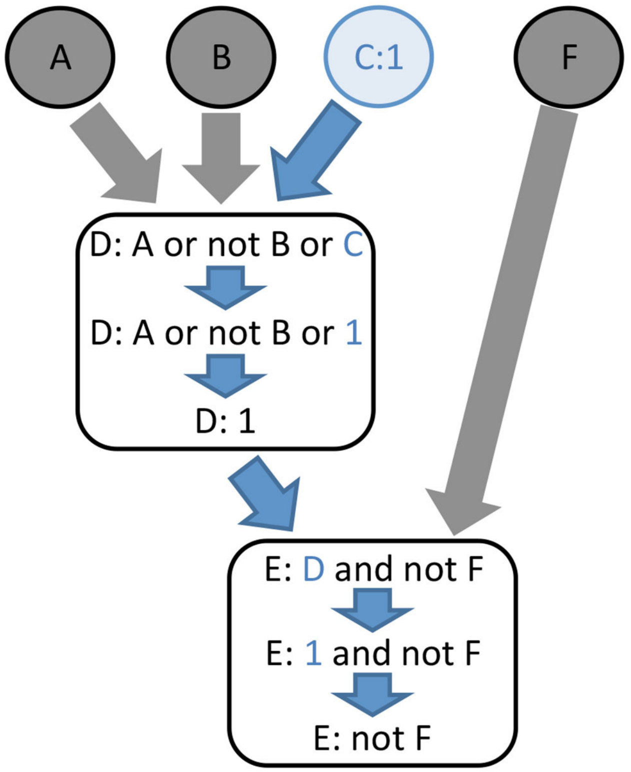


FIGURE 3 | The principle of logical value propagation analysis is illustrated with a simple example involving two core nodes, D and E, and four input nodes, A, B, C, and F. The value 1 is assigned to the node C and then propagated through the model. The assignment C:1 implies the evaluation of D to 1. Consequently, the function assigned to node E becomes "not F." In other words, assigning the value 1 to node C activates node D independently from the value of its other inputs, while node E becomes completely dependent on the value of node F.

Value Propagation Enables the Evaluation of the Impact of a Given Cellular Environment on Model Dynamics

The core idea of *value propagation* is presented in **Figure 3**. Given a set of logical rules and a cellular context, an iterative algorithm enables the computation of the dynamical consequences of the cellular context on all the components of the model.

First, the cellular context is formalized by assigning constant values to some components of the model. Next, we apply a recent model reduction technique reported by Saadatpour et al. (2013). Briefly, for each constant node, the corresponding value is inserted into the logical rule associated with each of its target nodes. Each logical rule is then simplified using Boolean algebra. If the rule simplifies to a constant, this fixed value is further propagated into the logical rules of downstream nodes. This process is iterated until no further propagation or simplification can be made on the logical rules of the model. In contrast with the approach of Saadatpour et al., which aims at producing a reduced model, we focus principally on the outcome of the propagation of fixed values.

The result of value propagation can be very informative by itself. Indeed, the resulting stabilized values provide insights into the impact of a given (single or multiple) perturbation on the model, revealing which elements are consequently constrained to become activated or inactivated, vs. which elements keep some degree of freedom. Furthermore, this method greatly eases the comparison of the impacts of different biological contexts on network dynamics by performing a differential analysis of the corresponding lists and target values of fixed components. This method has multiple advantages when applied to complex networks, as it can be used efficiently on models with large numbers of components. It further simplifies the computation of attractors (stable states or even simple or complex cycles). Interestingly, Saadatpour and collaborators showed that this method conserves the stable states and complex attractors under the fully asynchronous updating assumption (Saadatpour et al., 2013).

This method was extended to multilevel models and implemented into the Java bioLQM library (Naldi, 2018). In this implementation, the fixed components are conserved during value propagation, enabling a direct comparison of the propagated effects of alternative perturbations.

The power of this approach is demonstrated on a concrete example in the following section.

APPLICATION: ASSESSING THE EFFECT OF CHECKPOINT BLOCKADE THERAPIES ON T CELL ACTIVATION

Biological Background

Over the last decades, immunotherapies have been the subject of intense studies and led to great advances in the field of cancer treatment. Through the years, it has then been recognized that T cells often display a reduced ability to eliminate cancer cells, and that expression of co-inhibitory receptors at their surface accounts for this compromised function. Receptors like Cytotoxic T-lymphocyte protein 4 (CTLA4, also known as CD152) (Walunas et al., 1994; Leach et al., 1996) and Programmed cell death protein 1 (PD-1, also known as PDCD1 or CD279) (Ishida et al., 1992) have been particularly studied in that context. Antibodies blocking the pathways downstream of these co-inhibitors (checkpoint blockade therapies) have become standard treatment for metastatic melanoma (Robert et al., 2011; Simpson et al., 2013) and other cancers (Ribas and Wolchok, 2018), including non-small cell lung cancer, renal cell carcinoma, Hodgkin's lymphoma, Merkel cell carcinoma and many others. The successes of these studies led to an increasing interest in T cell co-inhibitory receptors.

Nevertheless, a clear understanding of the mechanisms at work inside T cells remains elusive. Therapies targeting CTLA4 or PD-1 show different immune adverse effects (June et al., 2017), while the corresponding intra-cellular mechanisms remain to be clarified. Moreover, a rationale for the educated development of new immunotherapies focusing on other receptors or combinations of receptors is clearly needed. Co-inhibitory receptors are legions at the surface of T cells (Brownlie and

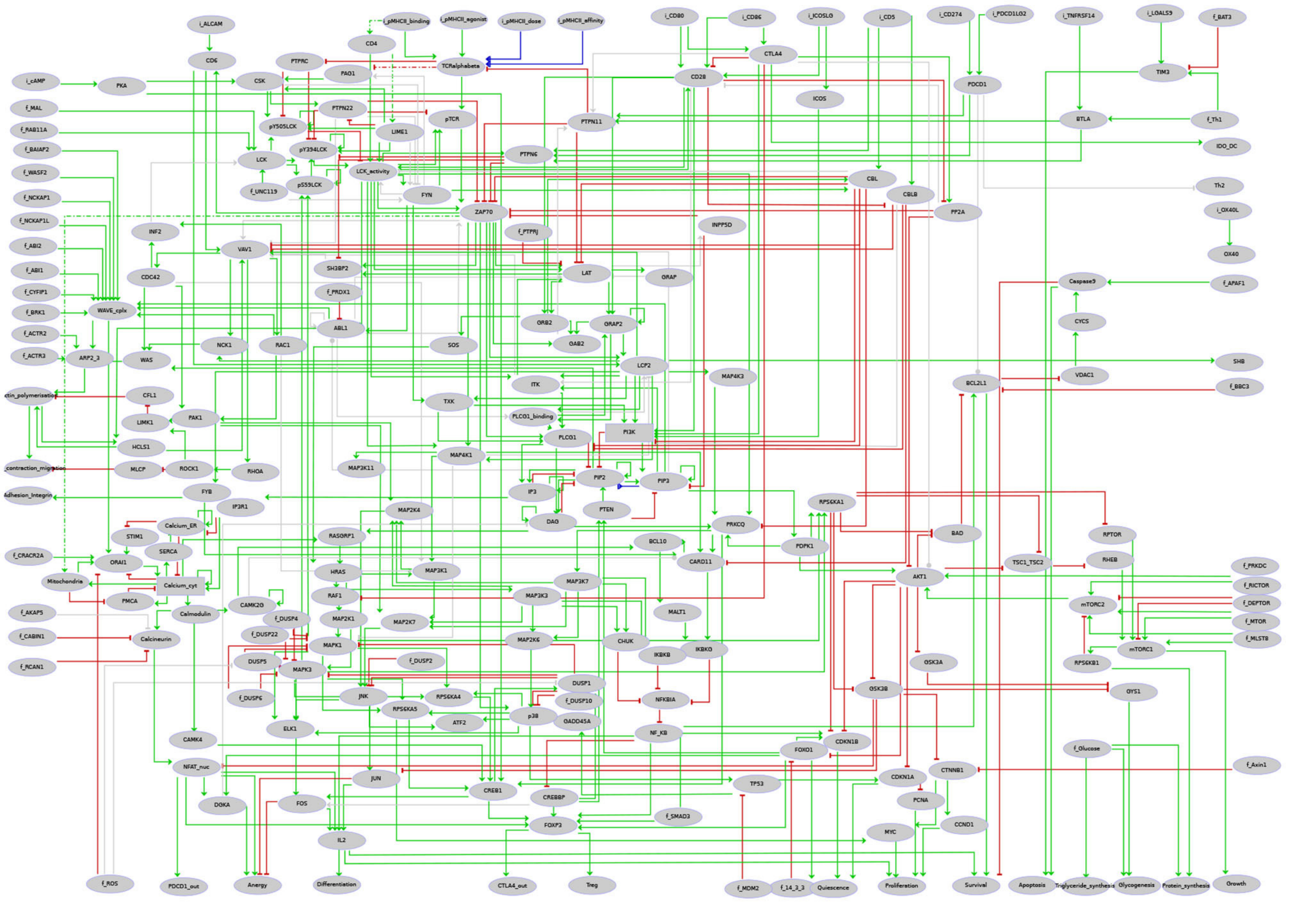


FIGURE 4 | Regulatory graph of the T cell activation model. The global layout is similar to the molecular map (cf. **Supplementary Figure 1** and **Supplementary File 1**), with ligands/receptors and proximal signaling at the top, and the nucleus-related events at the bottom of the graph. In between, the model encompasses interconnected pathways and signaling cascades related to cytoskeleton remodeling, the MAPK network, calcium fluxes, metabolic shifts, and NF-κB, to name a few. Boolean components are denoted by ellipsoids whereas rectangles denote ternary components. Green arcs denote activation events, red blunt arcs denotes inhibitions, while blue arcs denote dual regulations. The gray arcs represent interactions created during the translation of the molecular map into the regulatory graph, but that are not yet integrated at the dynamical level (i.e., not taken into account in the logical rule).

Zamoyska, 2013) and biology of T cell activation or tolerance involves activation or repression of highly interconnected and complicated pathways (Baumeister et al., 2016).

Given the central role of T cells in many medical contexts, several mathematical frameworks have been applied to model T cell activation. Recent examples include rule-based approaches (Chylek et al., 2014), ordinary differential equations (Perley et al., 2014), and logical models (Oyeyemi et al., 2015; Rodriguez-Jeorge et al., 2019; Sanchez-Villanueva et al., 2019), considering different biomedical contexts as diverse as HIV infection or neonate vaccination. To our knowledge, none of them specifically focused on the impact of co-inhibitory receptors on T cell activation or tolerance.

In this study, we applied the logical framework to integrate current data on CTLA4 and PD-1 pathways and assess their impact on T cell activation. Our goal was triple. First, we wanted to create a comprehensive model building upon extensive knowledge encoded into a molecular map (see next section). Second, using model verification and a specific unit test suite, we aimed to firmly anchor the model at both the global and local scale into the collected biological knowledge. Third, using value propagation, we aimed to provide a tool for the comparative analysis of intra-cellular consequences when targeting CTLA4 vs. PD-1 T-cell co-receptors.

Comprehensive Molecular Mapping of T Cell Activation Network

Prior to mathematical modeling, knowledge about biological entities involved in T Cell activation was collected from available pathway databases, including Reactome (Fabregat et al., 2016), PantherDB (Mi et al., 2013), ACSN (Kuperstein et al., 2015), and WikiPathways (Slenter et al., 2018). Moreover, the scientific literature indexed in the PubMed database was further explored and carefully curated. Using the software CellDesigner (version 4.3.1) (Funahashi et al., 2008), this knowledge was encoded in a molecular map describing reactions between biological entities (either proteins, RNAs, genes, complexes, or metabolites). Each biological entity included in the map was annotated with a series of standard identifiers, including UniProtKB accession number, recommended and alternative names, gene name and synonyms, and cross-references to unique HGNC identifiers and approved symbols. The annotations also reference relevant scientific articles, including PubMed identifier, first and last authors, year of publication, and a list of observations extracted from these publications.

Our T cell activation map currently encompasses 726 biological entities, in different states (active/inactive, with or without post-translational modifications), and 539 reactions involving these entities (**Supplementary Figure 1** and **Supplementary File 1**). Globally, the map currently integrates information from 123 scientific articles, which are cited in the annotations of the entities and reactions of the map.

Logical Modeling of T Cell Activation

Using the logical modeling software *GINsim* (version 3.0.0b) (Naldi et al., 2018a), we then manually derived a regulatory graph encompassing 216 nodes and 451 arcs (**Figure 4**) from the content of the molecular map. One by one, biological entities represented in the molecular map were re-created as components of the logical model. In most of the cases, the representation of entities having different states was further compressed into a single component summarizing their activity in the TCR signaling cascade. Furthermore, to obtain a dynamical logical model, a specific logical rule must be assigned to each node. In many cases, this can be achieved rather easily based on published data. For more complex situations, a default generic logical rule was initially considered, where all activators are needed for the activation of a component (using the AND operator) and where only one inhibitor is sufficient to repress it (using the OR and NOT operators), which served as a basis for further rule refinement. In some cases, however, in particular when a component is the target of various regulatory interactions or when metabolites are involved, finding direct support for a specific rule may be tricky or impossible. Hence, the delineation of consistent logical rules for a complex model is often the result of an iterative process, starting with generic rules and progressively correcting them based on the results of various analyses.

Hereafter, we demonstrate how we can take advantage of the methods presented in the previous sections to ease rule refinement by model verification. We first defined a series of properties expected for the model (see examples in **Table 1**). Next, stable states and/or trap spaces were computed and automatically compared with these properties (cf. first Jupyter notebook provided on the model web page at http://ginsim.org/model/tcell-checkpoint-inhibitors-tcla4-pd1). After some iterative runs of the notebook, manual refinements lead us to a set of rules complying with all the tests.

For example, the Endoplasmic Reticulum (ER) serves as a reservoir for calcium ions. This reservoir can be emptied through activation of the Inositol 1,4,5-trisphosphate receptor (IP3R1). When empty, this reservoir can be filled through activation of the Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA) pumps. A default logical rule for a node representing the presence of this Calcium quantity (Calcium_ER) is then "SERCA AND NOT IP3R1." To check the behavior of the corresponding logical sub-model, we defined a test checking whether whenever Calcium_ER was evaluated to TRUE, SERCA was evaluated to FALSE (see test "test_calc_tp_rest_ER1_SERCA0"). However, consecutive model verification failed, allowing us to notice that the default rule implied that SERCA should be always TRUE for Calcium_ER to be TRUE. The rule was then corrected to take into account the fact that Calcium_ER should stay TRUE whenever it would reach this value in absence of IP3R1.

TABLE 1 | Global specifications used to assess the T cell activation model and example of local specifications for the calcium signaling module (Cf. **Figure 2**).

Context	Expected behavior (active or inactive components)
Global specifications	
No stimulation	Active: Quiescence, Glycogenesis
Non-optimal stimulation (antagonist or low/high dosage/affinity)	Active: Anergy (DGKA), Quiescence, Glycogenesis
Optimal stimulation	Active: Differentiation, IL2, Proliferation, Growth, Actin polymerization, Actin contraction
CTLA4 stimulation	Active: Quiescence
PD-1 stimulation	Active: Quiescence
Local specifications for the calcium module	
Absence of IP3R stimulation, Calcium in ER	Active: Calcium_ER; Inactive: IP3R1
Absence of IP3R stimulation, Calcium in the cytoplasm	Active: Calcium_ER; Inactive: IP3R1
IP3R stimulation, Calcium in ER	Active: IP3R1, Calcineurin; Inactive: Calcium_ER

After verification, named components should have a value of 0 if specified as inactive, while active components should have a value of 1 or 2. Verification of local specifications requires the extraction of a sub-model from the global model. These verifications and literature references are detailed in the companion CoLoMoTo notebooks. ER: Endoplasmic Reticulum.

In the first Jupyter notebook provided as **Supplementary Material**, we include all the code enabling the verification of our final model, which encompasses 36 unit tests split in four test suites. On a MacBook Pro using macOS 10.13 High Sierra, with a 2.3 GHz Intel Core i7 and 16GB 1600 MHz DDR3, all the tests were run in 87s.

The four test suites cover the most complex parts of the model, some of them particularly difficult to define. These suites use sub-models, whose delineation was guided by known pathways and practical knowledge gained by the modeler during the assembly of the molecular map. The Calcium module test suite covers a sub-model related to the fluxes of Calcium ions between different cellular compartments, namely the endoplasmic reticulum, the cytoplasm, and the extracellular region. The LCK module test suite is centered on the Tyrosine-protein kinase Lck (LCK). This kinase is known to have multiple sites of phosphorylation, whose

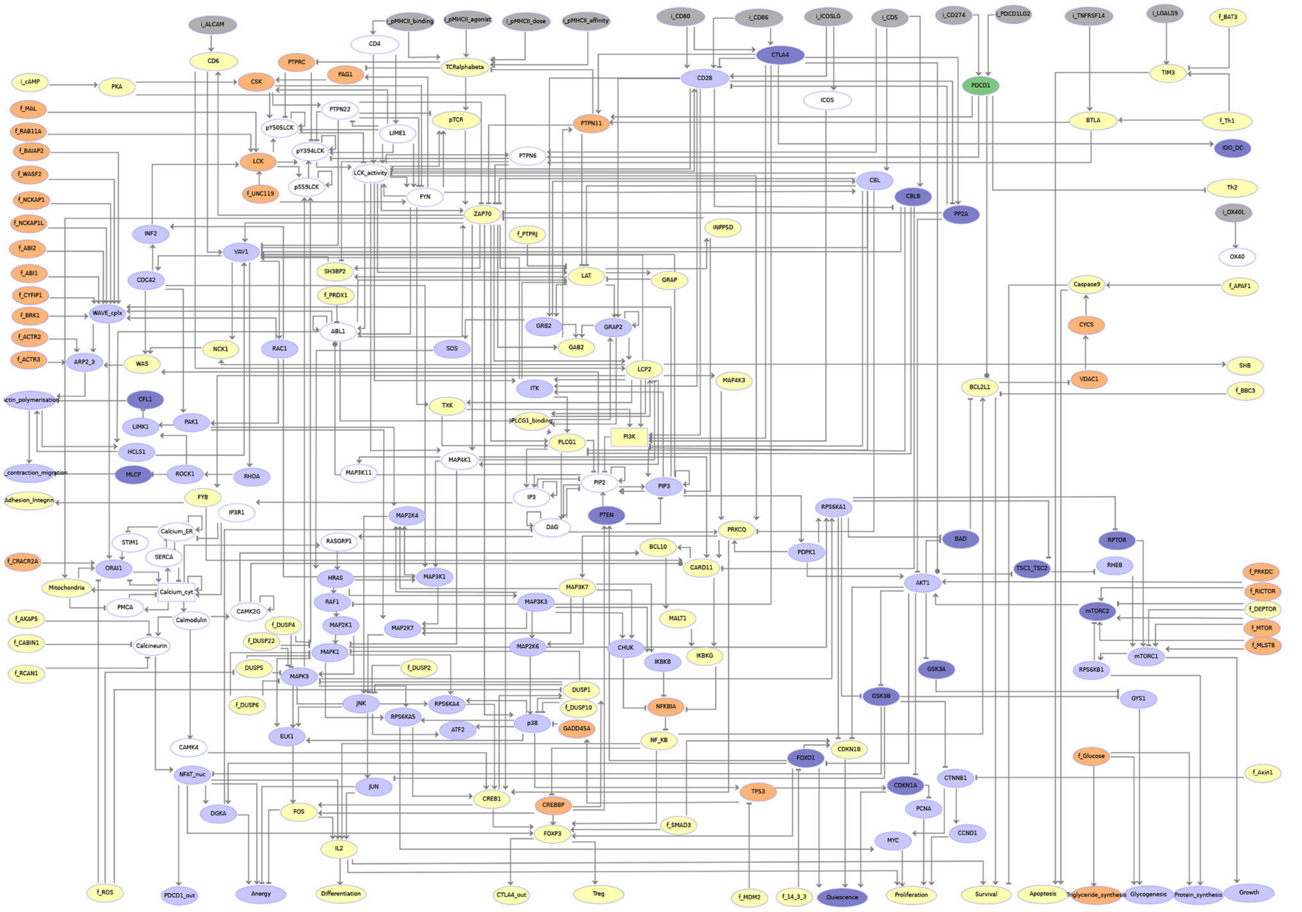


FIGURE 5 | Visualization of the results of the propagation analyses for CTLA4 vs. PD-1 activation. Gray nodes correspond to inputs. Nodes in yellow are frozen OFF upon any of CTLA4 or PD-1 (PDCD1) activation. Nodes in orange are frozen ON (i.e., with level 1 or 2) for each of these conditions. Nodes in light blue are frozen OFF only for CTLA4 activation (i.e., they remain free upon PD-1 activation). Nodes in dark blue are frozen ON (i.e., with level 1 or 2) only upon CTLA4 activation (i.e., they remain free upon PD-1 activation). Upon PD-1 activation, the corresponding node (PDCD1) is the only one that gets specifically frozen (ON, shown in dark green). Nodes in white remain free for both conditions.

collective status determines the tridimensional conformation and thus the activity of the enzyme (Ventimiglia and Alonso, 2013). The Cytoskeleton module test suite covers the cytoskeleton remodeling events occurring during T cell activation, and has strong connections with the Calcium sub-model. Finally, the Anergy/activation/differentiation module covers a less documented module encompassing the nucleus compartment and gene transcription.

Comparison of the Impacts of CTLA4 and PD-1 Co-Inhibitory Receptors Through Value Percolation

Based on the model described in the preceding section, a comparative propagation analysis was performed to visualize the respective effects of CTLA4 and PD-1 receptor activation on model dynamics. **Figure 5** displays the value propagation for each condition on a single regulatory graph, using a color code to distinguish the different situations (component inhibition/activation in one or both conditions). The value propagation for the two conditions are further shown separately in the second companion notebook (available at http://ginsim.org/model/tcell-checkpoint-inhibitors-tcla4-pd1). This analysis reveals that the activation of the CTLA4 receptor impacts most pathways of the model, impeding in particular the remodeling of the cytoskeleton and the metabolic switch associated with bona fide T cell activation. In contrast, the activation of the PD-1 receptor leads to more limited effects, predominantly freezing the components of the NF-κB pathway.

A more refined comparative analysis of value propagation from these two receptor activations entails the observation that the set of nodes frozen by the propagation of PD-1 activation is completely included inside the set of nodes frozen by the propagation of CTLA4 activation (see **Table 2**). Furthermore, the values of the components frozen in both propagation studies are the same. Interestingly, a set of nodes related to calcium influx from and to the endoplasmic reticulum remain unfixed by any of the propagation analyses. This could be an artfact of the positive feedback loops added on the nodes representing the Calcium ion levels in different compartments and would need to be further investigated. A more detailed biological interpretation of these results is proposed in the following section.

TABLE 2 | Quantification of the model nodes impacted by the propagation of CTLA4 or PD-1 persistent activation.

Impact of value propagation	CTLA4 ON	PD-1 ON	Intersection
Frozen inactive nodes	105	47	47
Frozen active nodes	28	13	12
Free nodes	29	102	28

After propagation, the nodes of the model can remain free (not fixed) or become frozen inactive (value 0) or frozen active (value 1, or potentially higher in the case of multilevel components). The model encompasses a total of 216 nodes, including 14 inputs and 40 nodes not affected by CTLA4 or PD-1 activation. Interestingly, PD-1 itself is the only single node specifically frozen by its activation (and not by the activation of CTLA4). Furthermore, each of 59 nodes affected by the two perturbations is frozen to the same value in both cases. Details of the computation method can be found in the companion CoLoMoTo notebooks.

CONCLUSIONS AND PROSPECTS

In this study, we have implemented and applied two complementary methods enabling a specification-oriented model building approach, thereby easing the delineation and analysis of highly complex logical models. In this respect, the building of a knowledge base, e.g., in terms of a molecular map, is an important first step. In the molecular map (provided as the **Supplementary File 1**), we have integrated the most relevant biological references available on T cell activation and inhibition pathways.

This map is clearly due to evolve, in particular thanks to the generation and analysis of novel high-throughput data (see e.g., the recent extensive analysis of the TCR signalosome by Voisinne et al., 2019). But any modification needs to be manually propagated to the dynamical model. To date, methods to derive proper dynamical models from such molecular maps are still in their infancy. In the particular case of the Boolean framework, only one automated approach has been recently proposed (Aghamiri et al., 2020). However, a limitation of this approach is the generation of generic logical rules based on static knowledge. Hence, the methods presented here could be used to advantageously refine these rules, taking into account additional biological knowledge about the behavior of the system under study.

We used the information gathered in our T cell activation map to build a dynamical logical model encompassing over 200 components and 450 interactions. For such a complex model, defining the logical rules in concordance with biological knowledge is a difficult and error-prone process, usually involving iterative trial simulations, where failures are identified to suggest potential improvements. Hence, listing comprehensive and consistent model specifications is a crucial step for model construction. These specifications can be revised as the modeler deepens his understanding of the biological processes under study. Noteworthy, such systematic testing procures a sense of confidence during the development process.

In the unit tests developed for our model, the definition of sub-models was guided by biological knowledge and pathway definitions, while relying partly on the modeler intuition. This step could be improved by community analyses of the regulatory graph to improve their definition.

Model checking techniques have been previously applied to assess model behavior through systematic cycles of model refinements (see e.g., Traynard et al. (2016) and reference therein). Model verification, as defined here, is a generalization of this approach, as it can rely on any available analysis as long as its result can be compared to an expected outcome. In our hands, in the course of model building, the unit testing approach, strongly anchored to available knowledge, proved to be very efficient to assess and improve model consistency with respect to a list of biological specifications, without the need of time-consuming and costly simulations. Implemented in the CoLoMoTo Interactive notebook framework (Naldi et al., 2018b), this approach enabled us to define a model recapitulating the most salient properties observed in response to T cell activation, including quiescence, anergy, and differentiation.

The use of model checking techniques could be further extended to assess the sensitivity of model behavior to the choice of specific logical rules. Such extension is hindered by the exponential increase of the number of possible logical rule, as the number of regulators increases. We would thus need a rationale to explore the space of logical rules. A first step in this direction can be found in Abou-Jaoudé and Monteiro (2019).

The approach presented here could also be improved by taking into account and tracking uncertainty during model conception (Thobe et al., 2018), or yet by taking advantage of computational repairing methods (Gebser et al., 2010) to identify more precisely remaining inconsistencies with biological data. Furthermore, other software engineering techniques, such as *code coverage*, could be borrowed to further improve model building and verification. Code coverage computes how much of a program's code is covered by unit tests. Similarly, one could design a method computing the fraction of the components of a model that is effectively covered by specifications.

Value propagation analysis of our large and complex regulatory graph proved to be biologically insightful. Indeed, this straightforward approach enabled us to clearly contrast the respective impacts of CTLA4 and PD-1 on T cell activation in our model, providing some rationale for their differential effects in current therapeutic studies. Indeed, anti-CTLA4 immunotherapies are known for their strong adverse effects related to autoimmunity and immunotoxicity (June et al., 2017). Anti-CTLA4 immunotherapies are currently combined with anti-PD-1 immunotherapy, known for its milder impact on the immune system.

Interestingly, the state of the node representing the Interleukin 2 (IL2) cytokine activation illustrates the differences of action of these receptors. Activation of the IL2 gene depends mainly on the activation of three transcription factors: the Nuclear Factor of Activated T cells (NFAT), the AP1 complex, and the Nuclear factor NF-kappa-B (NF-κB) (Smith-Garvin et al., 2009). When NFAT and AP1 are both active, they form a complex and together bind a regulatory region of the IL2 gene. In absence of AP1, NFAT induces a different program leading to cellular anergy (Macian, 2005; Smith-Garvin et al., 2009): activation of Diacylglycerol Kinase (DGK) prevents DAG-mediated activation of RasGRP1, which regulates the threshold for T cell activation (Roose et al., 2007; Das et al., 2009).

Our comparative propagation analysis reveals that while the activation of the CTLA4 receptor leads to a general inactivation of the three transcription factors regulating IL2 production, activation of the PD-1 receptor leads only to the inactivation of NF-κB and FOS (a member of the AP1 complex), thereby preventing the formation of the NFAT/AP1 complex, but enabling the activation of DGK. This observation is consistent with the proposal to target DGK isoforms as a complement of checkpoint immunotherapy (Riese et al., 2016; Jung et al., 2018).

As a next step, new co-inhibitory receptors recently under study, such as the Hepatitis A virus cellular receptor 2 (also known as TIM3) or the Lymphocyte activation gene 3 protein (LAG-3) (Anderson et al., 2016), could be easily added to the model described here, provided sufficient information could be gathered regarding their interacting partners. Applying propagation analysis in this context would be greatly insightful for future therapy developments.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

CH developed the T cell signaling molecular map and model under the supervision of MT-C and DT. CH and AN implemented the computational methods and applied them to the T cell model under the supervision of AN and DT. All co-authors contributed to the redaction of the manuscript and endorse its content. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a grant from the French Plan Cancer (project SYSTAIM, 2015–2019).

ACKNOWLEDGMENTS

We warmly thank Bernard Malissen, Romain Roncagalli, and Guillaume Voisinne for their thoughtful guidance during the construction of the T cell signaling model.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.558606/full#supplementary-material

Supplementary Figure 1 | Molecular map in good resolution or vectorial format Caption: This molecular map describes biological entities and reactions implicated in the process leading to activation of CD4+ T cells in humans/mouse. The cytoplasmic membrane and attached receptor proteins are placed at the top, while the cell nucleus is located at the bottom. Reactions represent current knowledge of various pathways related to cytoskeletal remodeling, calcium fluxes, metabolism, cell cycle or IL2 production, and are encoded using the CellDesigner software (Funahashi et al., 2008) (the CellDesigner file is provided as **Supplementary File 1**).

Supplementary File 1 | Molecular map in CellDesigner/XML format (Funahashi et al., 2008).

REFERENCES

Abou-Jaoudé, W., and Monteiro, P. T. (2019). On logical bifurcation diagrams. *J. Theor. Biol.* 466, 39–63. doi: 10.1016/j.jtbi.2019.01.008

Abou-Jaoudé, W., Monteiro, P. T., Naldi, A., Grandclaudon, M., Soumelis, V., Chaouiya, C., et al. (2014). Model checking to assess T-helper cell plasticity. *Front. Bioeng. Biotechnol.* 2:86. doi: 10.3389/fbioe.2014.00086

Aghamiri, S. S., Singh, V., Naldi, A., Helikar, T., Soliman, S., and Niarakis, A. (2020). Automated inference of Boolean models from molecular interaction maps using CaSQ. *Bioinformatics* 1–13. doi: 10.1093/bioinformatics/btaa484. Available online at: https://academic.oup.com/bioinformatics/advance-article

Anderson, A. C., Joller, N., and Kuchroo, V. K. (2016). Lag-3, Tim-3, and TIGIT: co-inhibitory receptors with specialized functions in immune regulation. *Immunity* 44, 989–1004. doi: 10.1016/j.immuni.2016.05.001

Baumeister, S. H., Freeman, G. J., Dranoff, G., and Sharpe, A. H. (2016). Coinhibitory pathways in immunotherapy for cancer. *Annu. Rev. Immunol.* 34, 539–573. doi: 10.1146/annurev-immunol-032414-112049

Boutillier, P., Maasha, M., Li, X., Medina-Abarca, H. F., Krivine, J., Feret, J., et al. (2018). The Kappa platform for rule-based modeling. *Bioinformatics* 34, i583–i592. doi: 10.1093/bioinformatics/bty272

Brownlie, R. J., and Zamoyska, R. (2013). T cell receptor signalling networks: branched, diversified and bounded. *Nat. Rev. Immunol.* 13, 257–269. doi: 10.1038/nri3403

Chakraborty, A. K. (2017). A perspective on the role of computational models in immunology. *Annu. Rev. Immunol.* 35, 403–439. doi: 10.1146/annurev-immunol-041015-055325

Chylek, L. A., Akimov, V., Dengjel, J., Rigbolt, K. T. G., Hu, B., Hlavacek, W. S., et al. (2014). Phosphorylation site dynamics of early T-cell receptor signaling. *PLoS ONE* 9:e104240. doi: 10.1371/journal.pone.0104240

Das, J., Ho, M., Zikherman, J., Govern, C., Yang, M., Weiss, A., et al. (2009). Digital signaling and hysteresis characterize Ras activation in lymphoid cells. *Cell* 136, 337–351. doi: 10.1016/j.cell.2008.11.051

Eftimie, R., Gillard, J. J., and Cantrell, D. A. (2016). Mathematical models for immunology: current state of the art and future research directions. *Bull. Math. Biol.* 78, 2091–2134. doi: 10.1007/s11538-016-0214-9

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., et al. (2016). The reactome pathway knowledgebase. *Nucleic Acids Res.* 44, D481–D487. doi: 10.1093/nar/gkv1351

Flobak, Å., Baudot, A., Remy, E., Thommesen, L., Thieffry, D., Kuiper, M., et al. (2015). Discovery of drug synergies in gastric cancer cells predicted by logical modeling. *PLoS Comput Biol.* 11:e1004426. doi: 10.1371/journal.pcbi.1004426

Funahashi, B. A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., and Kitano, H. (2008). A versatile modeling tool for biochemical networks. *Proc. IEEE* 96, 1254–1265. doi: 10.1109/JPROC.2008.925458

Gebser, M., Guziolowski, C., Ivanchev, M., Schaub, T., Siegel, A., Veber, P., et al. (2010). "Repair and prediction (under inconsistency) in large biological networks with answer set programming," in *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010* (Toronto, ON: AAAI Press).

Grieco, L., Calzone, L., Bernard-Pierrot, I., Radvanyi, F., Kahn-Perlès, B., and Thieffry, D. (2013). Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput. Biol.* 9:e1003286. doi: 10.1371/journal.pcbi.1003286

Hoops, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., et al. (2006). COPASI - A COmplex PAthway SImulator. *Bioinformatics* 22, 3067–3074. doi: 10.1093/bioinformatics/btl485

Ishida, Y., Agata, Y., Shibahara, K., and Honjo, T. (1992). Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* 11, 3887–3895.

June, C. H., Warshauer, J. T., and Bluestone, J. A. (2017). Is autoimmunity the Achilles' heel of cancer immunotherapy? *Nat. Med.* 23, 540–547. doi: 10.1038/nm.4321

Jung, I. Y., Kim, Y. Y., Yu, H. S., Lee, M., Kim, S., and Lee, J. (2018). CRISPR/Cas9-mediated knockout of DGK improves antitumor activities of human T cells. *Cancer Res.* 78, 4692–4703. doi: 10.1158/0008-5472

Kaufman, M., Andris, F., and Leo, O. (1999). A logical analysis of T cell activation and anergy. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3894–3899. doi: 10.1073/pnas.96.7.3894

Kaufman, M., Urbain, J., and Thomas, R. (1985). Towards a logical analysis of the immune response. *J. Theor. Biol.* 114, 527–561. doi: 10.1016/s0022-5193(85)80042-4

Klarner, H., Siebert, H., Nee, S., and Heinitz, F. (2018). Basins of attraction, commitment sets and phenotypes of Boolean networks. *IEEE ACM Trans. Comput. Biol. Bioinformatics* 17, 1115–1124. doi: 10.1109/TCBB.2018.2879097

Kuperstein, I., Bonnet, E., Nguyen, H.-A., Cohen, D., Viara, E., Grieco, L.,et al. (2015). Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis* 4:e160. doi: 10.1038/oncsis.2015.19

Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 16, 146–158. doi: 10.1038/nrg3885

Leach, D. R., Krummel, M. F., and Allison, J. P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. *Science* 271, 1734–1736. doi: 10.1126/science.271.5256.1734

Lopez, C. F., Muhlich, J. L., Bachman, J. A., and Sorger, P. K. (2013). Programming biological models in Python using PySB. *Mol. Syst. Biol.* 9, 1–19. doi: 10.1038/msb.2013.1

Macian, F. (2005). NFAT proteins: key regulators of T-cell development and function. *Nat. Rev. Immunol.* 5, 472–484. doi: 10.1038/nri1632

Mi, H., Muruganujan, A., and Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, 377–386. doi: 10.1093/nar/gks1118

Miskov-Zivanov, N., Zuliani, P., Wang, Q., Clarke, E. M., and Faeder, J. R. (2016). "High-level modeling and verification of cellular signaling," in *2016 IEEE International High Level Design Validation and Test Workshop, HLDVT 2016* (Santa Cruz, CA), 162–169.

Monteiro, P. T., and Chaouiya, C. (2012). "Efficient verification for logical models of regulatory networks," in *6th International Conference on Practical Applications of Computational Biology & Bioinformatics* (Berlin; Heidelberg: Springer), Vol. 154, 259–267.

Myers, G. J. (1979). *The Art of Software Testing, 1st Edn*. New Jersey, NJ: Wiley Publishing.

Naldi, A. (2018). BioLQM: a Java Toolkit for the manipulation and conversion of logical qualitative models of biological networks. *Front. Physiol.* 9:1605. doi: 10.3389/fphys.2018.01605

Naldi, A., Hernandez, C., Abou-Jaoudé, W., Monteiro, P. T., Chaouiya, C., and Thieffry, D. (2018a). Logical modeling and analysis of cellular regulatory networks with GINsim 3.0. *Front. Physiol.* 9:646. doi: 10.3389/fphys.2018.00646

Naldi, A., Hernandez, C., Levy, N., Stoll, G., Monteiro, P. T., Chaouiya, C., et al. (2018b). The CoLoMoTo interactive notebook: accessible and reproducible computational analyses for qualitative biological networks. *Front. Physiol.* 9:680. doi: 10.3389/fphys.2018.00680

Oyeyemi, O. J., Davies, O., Robertson, D. L., and Schwartz, J. M. (2015). A logical model of HIV-1 interactions with the T-cell activation signalling pathway. *Bioinformatics* 31, 1075–1083. doi: 10.1093/bioinformatics/btu787

Perley, J., Mikolajczak, J., Buzzard, G., Harrison, M., and Rundell, A. (2014). Resolving early signaling events in T-Cell activation leading to IL-2 and FOXP3 transcription. *Processes* 2, 867–900. doi: 10.3390/pr2040867

Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359, 1350–1355. doi: 10.1126/science.aar4060

Riese, M. J., Moon, E. K., Johnson, B. D., and Albelda, S. M. (2016). Diacylglycerol kinases (DGKs): novel targets for improving T cell activity in cancer. *Front. Cell Dev. Biol.* 4:108. doi: 10.3389/fcell.2016.00108

Robert, C., Miller, W. H., Francis, S., Chen, T.-T., Ibrahim, R., Hoos, A., et al. (2011). Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* 364, 2517–2526. doi: 10.1056/nejmoa1104621

Rodriguez-Jeorge, O., Kempis-Calanis, L., Abou-Jaoude, W., Hernandez, C., Ramirez-Pliego, O., Thomas-Chollier, M., et al. (2019). Cooperation between T cell receptor and Toll-like receptor 5 signaling for CD4+ T cell activation. *Sci. Signal.* 12:eaar3641. doi: 10.1126/scisignal.aar3641

Roose, J. P., Mollenauer, M., Ho, M., Kurosaki, T., and Weiss, A. (2007). Unusual interplay of two types of Ras activators, RasGRP and SOS, establishes sensitive and robust Ras activation in lymphocytes. *Mol. Cell. Biol.* 27, 2732–2745. doi: 10.1128/MCB.01882-06

Saadatpour, A., Albert, R., and Reluga, T. C. (2013). A reduction method for Boolean network models proven to conserve attractors. *SIAM J. Appl. Dynam. Syst.* 12, 1997–2011. doi: 10.1137/13090537x

Sanchez-Villanueva, J., Rodr-guez-Jorge, O., Ram-rez-Pliego, O., Rosas-Salgado, G., Abou-Jaoude, W. C. H., Naldi, A., et al. (2019). Contribution of ROS and metabolic status to neonatal and adult CD8+T cell activation. *PLoS ONE* 14:e0226388. doi: 10.1371/journal.pone.0226388

Sarma, G. P., Jacobs, T. W., Watts, M. D., Vahid Ghayoomie, S., Larson, S. D., and Gerkin, R. C. (2016). Unit testing, model validation, and biological simulation [version 1; referees: 2 approved, 1 approved with reservations]. *F1000Res.* 5, 1–17. doi: 10.12688/F1000RESEARCH.9315.1

Simpson, T. R., Li, F., Montalvo-Ortiz, W., Sepulveda, M. A., Bergerhoff, K., Arce, F., et al. (2013). Fc-dependent depletion of tumor-infiltrating regulatory T cells co-defines the efficacy of anti-CTLA-4 therapy against melanoma. *J. Exp. Med.* 210, 1695–1710. doi: 10.1084/jem.20130579

Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., et al. (2018). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids Res.* 46, D661–D667. doi: 10.1093/nar/gkx1064

Smith-Garvin, J. E., Koretzky, G. A., and Jordan, M. S. (2009). T cell activation. *Annu. Rev. Immunol.* 27, 591–619. doi: 10.1146/annurev.immunol.021908.132706

Thobe, K., Kuznia, C., Sers, C., and Siebert, H. (2018). Evaluating uncertainty in signaling networks using logical modeling. *Front. Physiol.* 9:1335. doi: 10.3389/fphys.2018.01335

Traynard, P., Fauré, A., Fages, F., and Thieffry, D. (2016). Logical model specification aided by model-checking techniques: application to the mammalian cell cycle regulation.

Bioinformatics 32, i772–i780. doi: 10.1093/bioinformatics/btw457

Ventimiglia, L. N., and Alonso, M. A. (2013). The role of membrane rafts in Lck transport, regulation and signalling in T-cells. *Biochem. J.* 454, 169–179. doi: 10.1042/BJ20130468

Voisinne, G., Kersse, K., Chaoui, K., Lu, L., Chaix, J., Zhang, L., et al. (2019). Quantitative interactomics in primary T cells unveils TCR signal diversification extent and dynamics. *Nat. Immunol.* 20, 1530–1541. doi: 10.1038/s41590-019-0489-8

Walunas, T. L., Lenschow, D. J., Bakker, C. Y., Linsley, P. S., Freeman, G. J., Green, J. M., et al. (1994). CTLA-4 can function as a negative regulator of T cell activation. *Immunity* 1, 405–413. doi: 10.1016/1074-7613(94)90071-x

Wang, Q., Miskov-Zivanov, N., Liu, B., Faeder, J. R., Lotze, M., and Clarke, E. M. (2016). "Formal modeling and analysis of pancreatic cancer microenvironment," in *Computational Methods in Systems Biology*, eds E. Bartocci, P. Lio, and N. Paoletti (Heidelberg: Springer International Publishing), 289–305.

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 16 October 2020
doi: 10.3389/fphys.2020.557721

Computational Modeling Analysis of Generation of Reactive Oxygen Species by Mitochondrial Assembled and Disintegrated Complex II

Nikolay I. Markevich[1]*, Lubov N. Markevich[2] and Jan B. Hoek[3]

[1] Institute of Theoretical and Experimental Biophysics RAS, Pushchino, Russia, [2] Institute of Cell Biophysics of RAS, Pushchino, Russia, [3] MitoCare Center for Mitochondrial Research, Department of Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, Philadelphia, PA, United States

Reactive oxygen species (ROS) function as critical mediators in a broad range of cellular signaling processes. The mitochondrial electron transport chain is one of the major contributors to ROS formation in most cells. Increasing evidence indicates that the respiratory Complex II (CII) can be the predominant ROS generator under certain conditions. A computational, mechanistic model of electron transfer and ROS formation in CII was developed in the present study to facilitate quantitative analysis of mitochondrial ROS production. The model was calibrated by fitting the computer simulated results to experimental data obtained on submitochondrial particles (SMP) prepared from bovine and rat heart mitochondria upon inhibition of the ubiquinone (Q)-binding site by atpenin A5 (AA5) and Complex III by myxothiazol, respectively. The model predicts that only reduced flavin adenine dinucleotide ($FADH_2$) in the unoccupied dicarboxylate state and flavin semiquinone radical ($FADH^•$) feature the experimentally observed bell-shaped dependence of the rate of ROS production on the succinate concentration upon inhibition of respiratory Complex III (CIII) or Q-binding site of CII, i.e., suppression of succinate-Q reductase (SQR) activity. The other redox centers of CII such as Fe-S clusters and Q-binding site have a hyperbolic dependence of ROS formation on the succinate concentration with very small maximal rate under any condition and cannot be considered as substantial ROS generators in CII. Computer simulation results show that CII disintegration (which results in dissociation of the hydrophilic SDHA/SDHB subunits from the inner membrane to the mitochondrial matrix) causes crucial changes in the kinetics of ROS production by CII that are qualitatively and quantitatively close to changes in the kinetics of ROS production by assembled CII upon inhibition of CIII or Q-binding site of CII. Thus, the main conclusions from the present computational modeling study are the following: (i) the impairment of the SQR activity of CII resulting from inhibition of CIII or Q-binding site of CII and (ii) CII disintegration causes a transition in the succinate-dependence of ROS production from a small-amplitude sigmoid (hyperbolic) shape, determined by Q-binding site or [3Fe-4S] cluster to a high-amplitude bell-shaped kinetics with a shift to small subsaturated concentrations of succinate, determined by the flavin site.

Keywords: complex II, reactive oxygen species, computational model, assembled, disintegrated

Edited by:
Denis Tsygankov,
Georgia Institute of Technology,
United States

Reviewed by:
Alexey Goltsov,
Abertay University, United Kingdom
Spyros K. Stamatelos,
Sanofi, United States

Correspondence:
Nikolay I. Markevich
markevich.nick@gmail.com

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: *30 April 2020*
Accepted: *31 August 2020*
Published: *16 October 2020*

Citation:
Markevich NI, Markevich LN and Hoek JB (2020) Computational Modeling Analysis of Generation of Reactive Oxygen Species by Mitochondrial Assembled and Disintegrated Complex II.
Front. Physiol. 11:557721.
doi: 10.3389/fphys.2020.557721

INTRODUCTION

Increasing interest in mitochondrial reactive oxygen species (ROS) production is caused by their crucial role not only in oxidative cellular damage and a development of various pathologies and aging but also in cell signaling that promote health by preventing a number of chronic diseases and extend lifespan (Ristow and Schmeisser, 2014). It was believed for a long time that complex I (CI) and respiratory Complex III (CIII) were the main producers of ROS in the respiratory chain (Turrens, 2003). The ability of respiratory complex II (CII) to generate ROS was debated, although there were experimental data (McLennan and Degli Esposti, 2000) allowing to conclude that CII can be a substantial source of ROS in mammalian mitochondria. It was found (McLennan and Degli Esposti, 2000) that CII is the predominant generator of ROS during prolonged respiration under uncoupled conditions, and CII appears to contribute to the basal production of ROS in cells.

The most clear evidence of CII of mammalian mitochondria to be a significant source of ROS under certain conditions was demonstrated in recent studies with inhibitors of CI, CII, and CIII (Quinlan et al., 2012; Siebels and Dröse, 2013; Grivennikova et al., 2017). It was found under condition when CI and CIII were inhibited in order to exclude ROS production from these complexes, CII could produce O_2^-/H_2O_2 in significant amounts, comparable and even exceeding ROS generated by CI and CIII, but only in the subsaturating range of succinate concentration. A dependence of the rate of ROS production on succinate concentration is bell-shaped with a maximum near 1,000 pmol/min mg prot at succinate concentration from approximately 50–500 μM in the experiments with both submitochondrial particles (SMP) and intact mitochondria (Quinlan et al., 2012; Siebels and Dröse, 2013; Grivennikova et al., 2017).

It was shown on SMP from bovine heart mitochondria (Siebels and Dröse, 2013) that excessive ROS production by CII at subsaturating succinate concentrations occurs due to a suppression of succinate-Q reductase (SQR) activity by different inhibitors such as atpenin A5 (AA5), blocker of ubiquinone (Q)-binding site, or stigmatellin, inhibitor of CIII. It was pointed out (Siebels and Dröse, 2013) that both inhibitors block SQR activity of CII: AA5 directly by blocking Q-binding site while the CIII inhibitor stigmatellin indirectly by decreasing the concentration of oxidized Q needed for the SQR activity. The suppression of SQR activity results in a strong succinate-dependent reduction of flavin adenine dinucleotide (FAD) and Fe-S redox centers located upper Q-binding site in CII that can induce significant ROS production by these centers at low succinate concentration.

While some authors (Quinlan et al., 2012) insist that only reduced $FADH_2$ in the unoccupied dicarboxylate binding site, and probably, flavin semiquinone radical ($FADH^{\bullet}$) should be a generator of ROS to account for their experimental data, others (Grivennikova et al., 2017) believe that only terminal [3Fe-4S] cluster is most suitable for the role of a redox center that forms O_2^-, judging by the experimentally observed midpoint redox potential of a hypothetical electron donor for $O_2^-$of about 40 mV.

It was found (Lemarie et al., 2011) that a direct impairment of the SQR activity of CII that induces very high ROS production can occur as a result of CII disintegration. CII is composed of hydrophilic FAD- and iron-sulfur-containing subunits (SDHA and SDHB, respectively) bound to a two-subunit (SDHC and SDHD) hydrophobic membrane anchor that contains cyt *b* and Q-binding site. It was shown (Lemarie et al., 2011) that the SQR activity of CII can be specifically impaired without affecting the succinate dehydrogenase (SDH) activity of this CII. This is achieved by the specific dissociation of the SDHA/SDHB subunits, which encompass the SDH activity, from the membrane-bound SDHC/SDHD complex that is required for the SQR activity. Such disintegration of CII can result from the pH decline or mitochondrial Ca^{2+} influx (Hwang et al., 2014), and depends on the diphosphatidylglycerol cardiolipin (Schwall et al., 2012; Hwang et al., 2014). Besides, a broad range of human diseases from cancers to neurodegeneration related to SDH malfunction have recently been linked to defective assembly factors (Moosavi et al., 2019).

It was found recently (Korge et al., 2017) that the excessive ROS production by CII with bell-shaped dependence on succinate concentration under suppression of the SQR activity of CII can result in autocatalytic mitochondrial permeability transition (MPT) due to efflux of succinate from mitochondria through the open MPT pore and activate ROS production at low succinate concentration with following activation of apoptosis or necrosis/necroptosis.

Despite of the important role of CII as a ROS generator and a sensor of apoptosis, mechanisms of ROS formation by this complex remain insufficiently understood. First of all, there is no consensus as to which of the redox centers of CII really form O_2^- and H_2O_2 with the bell-shaped dependence of the rates of ROS production on the succinate concentration observed experimentally (Quinlan et al., 2012; Siebels and Dröse, 2013; Grivennikova et al., 2017). Besides, it is not clear what changes occur in the kinetics of ROS formation by different sites of CII at CII transition from the assembled to disintegrated state. In order to answer some of these questions, a computational mechanistic model of electron transfer and O_2^-/H_2O_2 formation at different sites of CII in the assembled and disintegrated states is developed in the present study which is a continuation of our previous theoretical studies of CII (Markevich et al., 2019). Previously, we studied a simplified model of electron transfer in the assembled state of CII only and without heme *b* as an electron carrier in order to account for qualitatively experimentally observed high-amplitude bell-shaped responses of ROS production in CII upon inhibition of CIII. In the present study, the model is significantly extended by including heme *b* and consideration of CII in both assembled and disintegrated states. In addition, the model has been calibrated by fitting the computer simulated results to experimental data obtained on SMP prepared from bovine heart mitochondria

---

Abbreviations: ROS, Reactive oxygen species; CII, Respiratory Complex II; CIII, Respiratory Complex III; FAD, Flavin adenine dinucleotide; SDH, Succinate dehydrogenase; Q, Ubiquinone; SQR, Succinate-Q reductase; SMP, Submitochondrial particles.

upon inhibition of Q-binding site by AA5 (Siebels and Dröse, 2013) and from rat heart mitochondria upon inhibition of Complex III by myxothiazol (Grivennikova et al., 2017).

Kinetics of ROS generation by each redox center able potentially to form O_2^-or H_2O_2 was analyzed using the developed model under different conditions (assembled and disintegrated states of CII, inhibition of CIII, and Q-binding site of CII) to account for available experimental data on ROS production and make predictions to be tested experimentally.

METHODS AND MODELS

Kinetic Models of Electron Transfer in Assembled and Disintegrated CII

Kinetic schemes of electron transfer and O_2^-/H_2O_2 production underlying a mechanistic computational model of CII in assembled and disintegrated states are presented in **Figure 1**. **Figure 1A** presents assembled CII, and **Figures 1B,C** – SDHA/SDHB and SDHC/SDHD subunits of disintegrated CII, respectively. Index "*d*" next to the names of redox centers and number reactions in **Figures 1B,C** means "disintegrated." These kinetic schemes include the following electron carriers: (a) $FADH_2$, (b) the sequence of iron-sulfur clusters: [2Fe-2S], [4Fe-4S], and [3Fe-4S], and (c) coenzyme Q. Electron transfer in CII takes into account both the mainstream electron pathway from succinate to Q (SQR activity) and bypass reactions resulting in O_2^-/H_2O_2 formation. These bypass reactions are marked in red in the kinetic schemes (**Figure 1**).

Electron pathway in kinetic schemes (A–C) can be described as follows.

After reduction in the reactions (1–3) in **Figure 1A** and (1d–3d) in **Figure 1B**, $FADH_2$ donates the first electron to the [2Fe-2S] cluster [reactions (6) in **Figure 1A** and (6d) in **Figure 1B**]. Then, the first electron through cluster [4F-4S] [reactions (7, 7d)] goes to the terminal cluster [3F-4S] [reactions (8, 8d) in **Figures 1A,B**, respectively]. In the assembled CII oxidized Q, binds to the Q-binding site [CII.Q site, reaction (10)], thus, the first electron can transfer from [3F-4S] cluster to Q-binding to CII either indirectly through cyt *b* [reactions (9, 11)] or directly in the reaction (13). Thus, transfer of the first electron from $FADH_2$ to the CII.Q results in generation of semiquinone anion ($CII.Q^-$) in the reactions (11, 13). Then, semiquinone anion binds proton H^+ forming protonated semiquinone radical CII.QH in the reaction (12).

Potential sites of ROS generation in this branch of the first electron transfer in **Figure 1A** are $FADH_2$, $CII.Q^-$, and [3Fe-4S]. $FADH_2$ can generate either H_2O_2 in the reaction (22) or superoxide in the reaction (23). Semiquinone, CII.Q, and [3Fe-4S] clusters generate superoxide in the reactions (25, 26). The same events occur in the matrix located subcomplex SDHA/B (**Figure 1B**) that is dissociated from the membrane-anchoring SDHC/D subcomplex. Difference between assembled and disintegrated subcomplex SDHA/B is only in downstream electron transfer. Oxidized $[3Fe\text{-}4S]^-$ cluster in disintegrated CII cannot donate electron to the cyt *b* and CII binding Q located in the membrane. In this case, $[3Fe\text{-}4S]^-$ donates electron to the oxygen only forming superoxide in the reaction (25d; **Figure 1B**).

By analogy, the second electron from the $FADH^\cdot$ radical transfers to the protonated semiquinone $CII.QH^\cdot$ in the reactions (14–18 and 21) with following formation and releases QH2 to the matrix in the reactions (18–21). In this branch of the second electron transfer, potential sites of superoxide formation are the $FADH^\cdot$ radical [reaction (24)] and $[3Fe\text{-}4S]^-$ [reaction (26)]. In addition, reaction (27) describes binding of AA5, inhibitor of CII, to the Q-binding site forming the inactive complex CII.AA5. In this reaction, AA5 competes with Q for binding to the Q-binding center, and thus competitively inhibits CII.

The entire reaction network of electron transfer and ROS production corresponding to kinetic schemes in **Figure 1** consists of 53 reactions and is described in detail in **Supplementary Table S1**. The values of midpoint redox potentials,

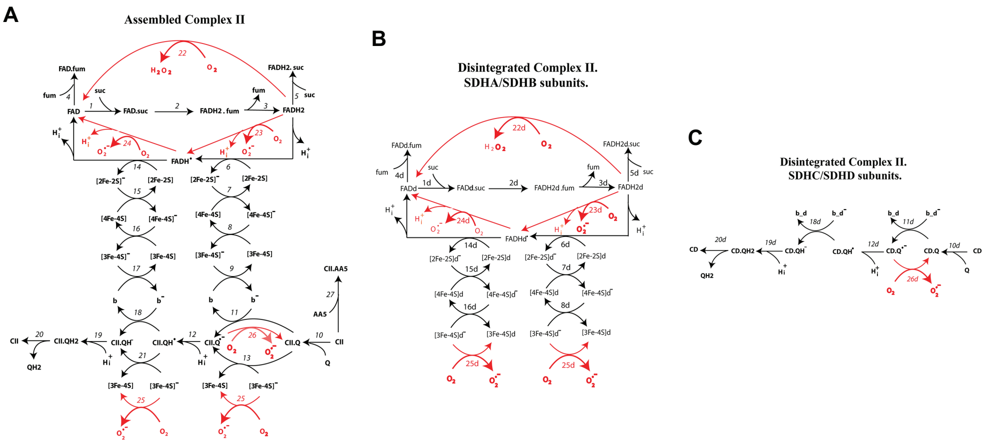


FIGURE 1 | Kinetic schemes of electron transfer and formation of superoxide anion, O_2^-, and hydrogen peroxide, H_2O_2, in assembled **(A)** and disintegrated **(B,C)** Complex II. Reactions of O_2^- and H_2O_2 formation are shown by red arrows. The detailed reaction network is presented in **Supplementary Tables S1, S2**.

rate constants, and concentrations of different electron carriers were taken from experimental data and presented in **Supplementary Table S2**.

Computational Model of Electron Transfer in Assembled and Disintegrated CII

A computational model consisting of 35 ordinary differential equations (ODEs) and 15 moiety conservation equations was derived from the reaction network using the law of mass action and Michaelis kinetics for all 53 kinetic processes. The model was implemented in DBSolve Optimum software available at http://insysbio.ru. The details of the mathematical model describing oxidized and reduced states of different carriers and electron flows through complex II are presented in **Supplementary Data**. The kinetic parameters used in the model are consistent with either measured or estimated values reported in the literature. We initially assumed the parameter values specified in the work (Markevich et al., 2019). They are presented in **Supplementary Table S2**. The simulated dependence of the rate of ROS production by CII on the succinate concentration demonstrated a good fit to the experimental data upon inhibition of Complex III. We subsequently fitted model simulation results to various experimental data, i.e., optimized some parameter values within experimental or estimated constraints to minimize the square deviation of the residuals, $\sum(\mathrm{Res}_i)^2$, where each residual (Res_i) is the difference between the experimental data point *i* and the value calculated by numerical computation of the model in the steady state for the given parameter set. Finding stationary solutions of the model, i.e., solving a system of algebraic equations, and the fitting procedure was performed by the DBSolve Optimum software in the Implicit Solver and Fitter options, respectively. **Supplementary Table S3** lists the values of the adjustable kinetic parameters.

Additionally, the model is presented in SBML format by separate file: 2019_cII_Final.xml as supporting information.

RESULTS AND DISCUSSION

Assembled CII

In order to investigate what redox centers of CII responsible for the experimentally observed high-amplitude bell-shaped dependence of the total rate of ROS generation by CII on the succinate concentration upon inhibition of CIII or Q-binding site (Quinlan et al., 2012; Siebels and Dröse, 2013; Grivennikova et al., 2017), i.e., suppression of the SQR activity of assembled CII as was pointed in Siebels and Dröse (2013), kinetics of ROS formation by each redox center able potentially to form O_2^- or H_2O_2 was analyzed computationally using mathematical modeling simulation of different inhibitory conditions.

Computer Simulated Inhibition of the Q-Binding Site of Assembled CII

Initially, in order to calibrate the developed model, computer simulated dependencies of the rates of ROS production on the AA5, inhibitor of the Q-binding site, and succinate concentration were fitted to experimental data on SMP prepared from bovine heart mitochondria (Siebels and Dröse, 2013). Computer simulated results and experimental data presented in **Figure 2** show a good fit for both dependencies of the H_2O_2 production rate on the AA5 concentration at the fixed succinate concentration (100 μM; **Figure 2A**) and on the varied succinate concentration at the fixed AA5 concentration (0.05 μM; **Figure 2B**). Fitting resulted in changes in values of some model parameters compare to the initial values including literature data. The new adjustable values of some parameters are presented in **Supplementary Table S3**. The most significant deviations of the adjusted values from the initial values are observed for the catalytic constants of the rate of ROS generation by each redox center and the catalytic constants of Q binding to the Q-binding center. Computer simulated time course of the total H_2O_2 production rate at changes in the AA5 concentration from 0 up to 0.05, 0.1, and 0.15 μM at fixed succinate concentration (100 μM; **Figure 2C**) confirms stationary fit computer simulated and experimental data presented in **Figure 2A**. Computer simulated dependencies of the stationary rate of QH2 production (succinate oxidation) on the succinate concentration at different fixed AA5 concentrations presented in **Figure 2D** show that the maximum reaction rate and the Michaelis constant decrease proportionally with increasing AA5 concentration which is typical for a competitive inhibitor.

Computer simulated dependence of stationary rates of ROS production by different redox centers of CII, namely: the flavin site, [3Fe-4S] cluster, and semiquinone, CII.Q^-, at the Q-binding site on the succinate concentration with varying degrees of inhibition of the Q-binding site by AA5 are presented in **Figure 3**. These computer simulation results predict that only reduced $FADH_2$ in the unoccupied dicarboxylate state (**Figures 3A,B**) and $FADH^\bullet$ (**Figure 3C**) have the experimentally observed [5] bell-shaped dependence of the rate of ROS production on the succinate concentration upon inhibition of the Q-binding site by AA5. Succinate-dependence of the rate of ROS formation at these sites (**Figures 3A–C**) as well as the total rate of ROS generation by CII (**Figure 3F**) changes from a very small amplitude in the basal state (AA5 = 0) to the intermediate- and high-amplitude bell-shaped kinetics with a shift to small succinate concentration upon an intermediate (AA5 = 0.05 μM) and strong (AA5 = 0.15 μM) inhibition of the Q-binding site, respectively.

It is proposed that $FADH_2$ can generate both hydrogen peroxide, H_2O_2, and superoxide, O_2^-, with the rates v_{22} and v_{23}, respectively (**Figure 1A**; **Supplementary Table S1**). The ratio of the catalytic constants of H_2O_2, and O_2^- formation k_{22} and k_{23} are unknown, one authors (Siebels and Dröse, 2013) found 75% H_2O_2 and 25% O_2^- in rat heart mitochondria while another (Grivennikova et al., 2017) found that bovine heart mitochondrial respiratory CII generates ROS, mostly as superoxide. Our fitting results predict close values for these constants: $k_{22} = 0.027\ \mu M^{-1}\cdot s^{-1}$ and $k_{23} = 0.019\ \mu M^{-1}\cdot s^{-1}$, so v_{22} (**Figure 3A**) and v_{23} (**Figure 3B**) look similar although we have to take into account that v_{23} is the rate of O_2^- production, so the contribution of v_{23} in the total rate of H_2O_2 production two times less than v_{22} because two molecules O_2^- give one

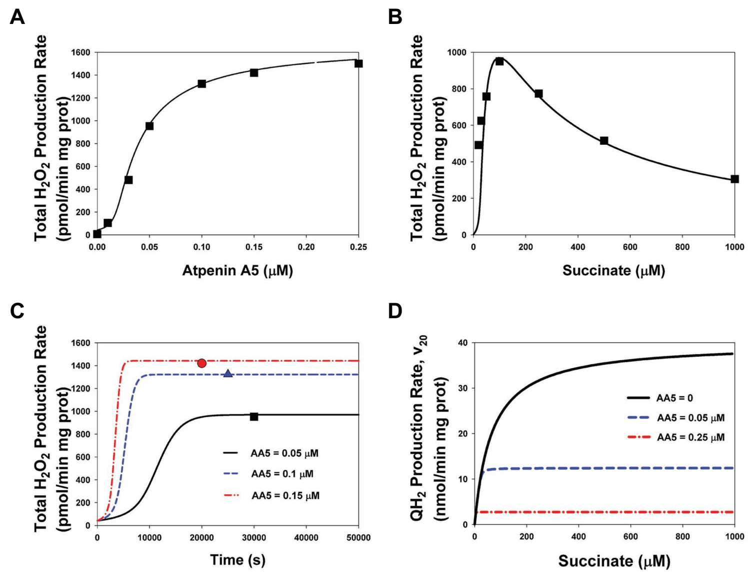


FIGURE 2 | The stationary rates of total H_2O_2 and QH_2 production by Complex II at varied concentrations of succinate and atpenin A5 (AA5). Computer simulation with the model of assembled respiratory Complex II (CII; lines) is compared to experimental data (symbols) on submitochondrial particles (SMP) prepared from bovine heart mitochondria (Siebels and Dröse, 2013). **(A,B)** The model simulated (solid line) and experimentally observed (black squares) dependence of the stationary rate of total H_2O_2 production on the AA5 concentration **(A)** (the concentration of succinate is equal to 100 μM) and the succinate concentration **(B)** (the AA5 concentration is equal to 0.05 μM). **(C)** Computer simulated time responses in the total rate of H_2O_2 production (lines) and its steady-state experimentally observed values [as shown in the section **(A)**] when the AA5 concentration changes from 0 to 0.05 (solid line and black squares), 0.1 (blue dashed line and triangles), and 0.15 (red dash-dot line and circles) μM. **(D)** The dependence of the computer simulated stationary rates of QH_2 production (succinate oxidation) on the succinate concentration at varied AA5 concentration that are shown in the plane: AA5 = 0 (black solid line), AA5 = 0.05 (blue dashed line), and 0.25 (red dash-dot line) μM. All computer simulations were made at the total concentration of CII, CIIt, of 235 μM and $k_{29} = 1\ s^{-1}$. The rest model parameters are presented in **Supplementary Tables S2, S3**.

molecule H_2O_2 in the process of subsequent dismutation (reaction 28) in **Supplementary Table S1**. The stationary total rate of H_2O_2 production by CII (vH_2O_{2tot}) was computed as the rate of H_2O_2 release from the mitochondrial matrix to cytosol that equal to the summary rate of H_2O_2 production by $FADH_2$ in assembled, v_{22}, and disintegrated, v_{22d}, states and dismutation of O_2^-, v_{28}, in the matrix at the steady state (see *Explicit functions* in Mathematical model in **Supplementary Data**).

The computer simulated dependence of stationary rates of ROS (O_2^-) production by [3Fe-4S] cluster, v_{25}, (**Figure 3D**) and semiquinone at the Q-binding site, v_{26}, (**Figure 3E**) on the succinate concentration show monotonic hyperbolic kinetics in the basal state (AA5 = 0) as well as upon an intermediate (AA5 = 0.05 μM) and strong (AA5 = 0.25 μM) inhibition of the Q-binding site by AA5. It should be pointed that AA5 increases the sensitivity of O_2^- production by [3Fe-4S] cluster, v_{25}, to succinate, i.e., decreases Michaelis constant, K_{m25}, and does not affect on the maximal rate v_{25}. On the contrary, AA5 decreases proportionally the maximal rate of O_2^- production by semiquinone at the Q-binding site, v_{26}, and K_{m26}.

Computer Simulated Inhibition of CIII

Just like in the work (Markevich et al., 2019), inhibition of CIII was simulated by decreasing the catalytic constant k_{29} of QH_2 oxidation in the mitochondrial inner membrane [reaction (29) in **Supplementary Tables S1, S2**]. **Figure 4** shows that model simulation of the stationary rate of H_2O_2 production at $k_{29} = 0.005\ s^{-1}$ ($k_{29} = 1\ s^{-1}$in the uninhibited state) good fit to experimental data on SMP from rat heart mitochondria upon inhibition of Complex III by myxothiazol (1.6 ìM; Grivennikova et al., 2017). It should be emphasized that values of all parameters including adjustable are the same for both modeling simulation presented in **Figures 2A,B, 4** except the total concentration of CII, CIIt, that is equal to 235 and 97 μM for **Figures 2, 4**, respectively. This fact strongly suggests that the calibration of the developed mathematical model is correct.

Computer simulation modeling results upon inhibition of CIII confirm suggested earlier hypothesis that the non-monotonic succinate-dependence of the rate of ROS generation in CII results from ROS formation at the flavin in the unoccupied dicarboxylate binding site (Quinlan et al., 2012). Computational modeling analysis shows that only $FADH_2$ and $FADH^{\bullet}$ have the experimentally observed bell-shaped dependence of the rate of ROS production on the succinate concentration upon inhibition of CIII and predict that inhibition of CIII has the same effects on the kinetics of ROS production by different CII redox centers (**Supplementary Figure S1**) as inhibition of the Q-binding site considered above (**Figure 3**). Such similar effect of inhibition of CIII and Q-binding site on the total ROS production by CII was pointed earlier as a result from suppression of SQR activity of CII (Siebels and Dröse, 2013). The stronger inhibition of CIII (a more decrease in k_{29}) results in the more amplitude

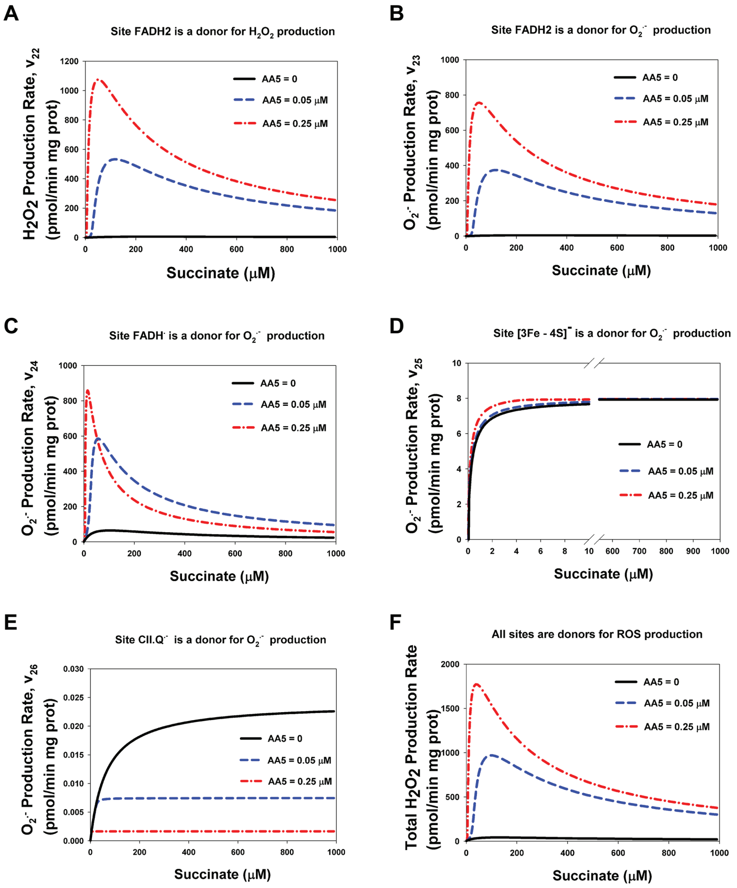


FIGURE 3 | Computer simulation of the effect of AA5 on the stationary rate of reactive oxygen species (ROS) production by different redox centers of the assembled Complex II. **(A–F)** Model simulated dependencies of the stationary rates of H_2O_2 **(A,F)** and O_2^- **(B–E)** production on the succinate concentration by different redox centers: flavin adenine dinucleotide ($FADH_2$) **(A,B)**, $FADH^\bullet$ **(C)**, [3Fe-4S] cluster **(D)**, and CII-binding semiquinone $CII.Q^-$ **(E)** at varied AA5 concentrations. **(F)** The total rate of H_2O_2 production by CII. Black solid, blue dashed, and red dash-dot lines correspond to AA5 = 0, 0.05, and 0.25 μM, respectively. All computer simulations were made at model parameter values as for **Figure 2**.

of the bell-shaped succinate-dependence of the rate of ROS production by the flavin site (**Supplementary Figures S1A–C**), and the total rate of ROS production by CII (**Supplementary Figure S1F**) and more shift of the maximal rate of ROS production to the small succinate concentration. The dependence of the stationary rate of ROS formation by [3Fe-4S] cluster, v_{25}, (**Supplementary Figure S1D**) and semiquinone at the Q-binding site, v_{26}, (**Supplementary Figure S1E**) on the succinate concentration shows hyperbolic kinetics in the basal state (k_{29} = 1 s^{-1}) as well as upon an intermediate (k_{29} = 0.1 s^{-1}) and strong (k_{29} = 0.01 s^{-1}) inhibition of CIII with very small amplitude. Thus, these computer simulation data predict that the effect of inhibition of CIII and the Q-binding site on ROS production by different sites of CII is very similar.

Bell-Shaped vs. Hyperbolic Kinetics of ROS Production and SQR Activity. What the Reason?

The hypothesis that a decrease in the rate of ROS production by flavin in the unoccupied dicarboxylate binding site at the high succinate concentration may be a reason for bell-shaped dependence of the total rate of ROS production by CII upon inhibition of CIII was first proposed in (Quinlan et al., 2012). It is really easy to understand a decrease in the concentration of ROS-producing site unoccupied $FADH_2$ due to succinate binding to $FADH_2$ in the reaction (5; **Figure 1A**; **Supplementary Table S1**). More difficult to understand hyperbolic dependence of the rate of ROS formation on the succinate concentration by [3Fe-4S] cluster and semiquinone at the Q-binding site as well as the concentration of these sites and other Fe-S clusters (**Supplementary Figure S2**) located downstream flavin. Moreover, experimentally observed catalytic activity of CII has also hyperbolic dependence on the succinate concentration (Quinlan et al., 2012), although at the first glance, bypass ROS production rate by these sites as well as the mainstream electron flow in CII, i.e., the rate of QH_2 production, v_{20}, should have non-monotonic dependence on the succinate concentration because of bell-shaped dependence of the concentration of $FADH_2$ and $FADH^\bullet$ (**Supplementary Figure S2**) that are substrate and product, respectively, in the mainstream and bypass reactions (6) and (23) (**Figure 1A**).

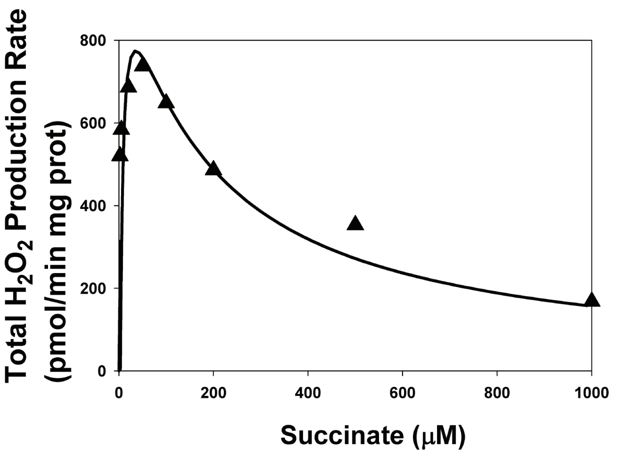


FIGURE 4 | The dependence of the stationary rate of total H_2O_2 production by CII on the succinate concentration upon inhibition of Complex III. Computer simulation (line) is compared to experimental data (black triangles) on SMP from rat heart mitochondria upon inhibition of Complex III by myxothiazol (1.6 ìM; Grivennikova et al., 2017). All computer simulations were made at model parameter values as for **Figure 2** except AA5 = 0, $k_{29} = 0.005$ s^{-1}, and the total concentration of CII, CIIt, equal to 97 μM.

Computer simulation results presented in **Figure 5** explain this situation. First, the computer simulated stationary rate v_{20} really has the hyperbolic dependence on the succinate concentration (**Figure 5A**) and decreases at decreasing k_{29}, i.e., decreasing QH_2 oxidation that results in a suppression of SQR activity. Second, **Figures 5B,C** show why the rates v_6 (that equal v_{20} in the steady state) and v_{23} (**Figure 1A**; **Supplementary Table S1**) that have one the same substrate and one the same product have qualitatively different kinetics, hyperbolic (v_6), and bell-shaped (v_{23}). **Figure 5B** shows that unidirectional mainstream electron flows in forward (from $FADH_2$ to oxidized [2Fe-2S] cluster), $v_{6\_forward}$, and reverse (from reduced $[2Fe\text{-}2S]^-$ cluster to $FADH^{\bullet}$), $v_{6\_reverse}$, direction (expression for them in the section **Supplementary Data** Mathematical model) really follow nonmonotonic $FADH_2$ and $FADH^{\bullet}$ concentration and have bell-shaped kinetics. However, the netto rate: $v_{6\_netto} = v_{6\_forward} - v_{6\_reverse}$ has hyperbolic kinetics due to common-mode inphase changes with a close amplitude in $v_{6\_forward}$ and $v_{6\_reverse}$ (**Figure 5B**). Values of these rates $v_{6\_forward}$ and $v_{6\_reverse}$ are so high compare to v_{6_netto} and close each other that curve for them are indistinguishable in **Figure 3B**. On the contrary, $v_{23\_reverse}$ is very small compare to $v_{23\_forward}$ (**Figure 5C**) due to a small concentration of the one product of this reaction, superoxide O_2^-. That is why the netto rate of superoxide O_2^- production by $FADH_2$, $v_{23\_netto} = v_{23\_forward} - v_{23\_reverse}$ follows the forward rate $v_{23\_forward}$ only and has bell-shaped kinetics.

The Effect of Limitation of Succinate Transport on ROS Production by CII

It is worth noting that the succinate concentration required for optimal (peak) ROS generation by CII upon inhibition of CIII or Q-binding site is different in different experiments and changes from approximately 50–500 μM in the experiments with both SMP and intact mitochondria (Quinlan et al., 2012; Siebels and Dröse, 2013; Grivennikova et al., 2017). It was proposed earlier (Grivennikova et al., 2017) that a high optimal succinate concentration up to 400–500 μM in intact mitochondria may be related to limitations of succinate permeability into the mitochondrial matrix, where succinate dehydrogenase active site is located. These limitations were simulated in the model by changes in the rate constant of succinate binding to FAD, k_1, because diffusion is one of the steps of succinate binding. Computational modeling results on changes in the rate constant k_1 presented in **Supplementary Figure S3** support, in part, this hypothesis. A decrease in k_1 results in a shift of the optimal succinate concentration to the high values of succinate with simultaneous decrease in the maximal rate of ROS production. These changes resemble the effect of inhibition of CIII and Q-binding site. Therefore, it seems most likely that different succinate optimal values of ROS production observed in different experiments are related to both different power of inhibition of CIII or Q-binding site in different experiments and transport limitations of succinate.

The Role of Heme *b* in the Electron Transfer in CII

It should point out that our preliminary computational modeling analysis of the simplified model of assembled CII without heme *b* as an electron carrier in the electron transfer pathway from succinate to Q at the Q-binding site (Markevich et al., 2019) predicts similar effects of inhibition of CIII on ROS production by CII, that is the high-amplitude bell-shaped dependence of the rate of ROS production by CII on the succinate concentration.

Computational analysis of the present extended model with heme *b*, i.e., included the thermodynamic cycle [3Fe-4S]↔heme *b*↔Q↔[3Fe-4S] [reactions (9, 11, and 13) for the first electron and reactions (17, 18, and 21) for the second electron transfer] that was experimentally studied in detail in (Anderson et al., 2014) shows that the value of midpoint potential of the heme *b*, $E_m(b)$, does not affect on the rates of QH_2 and ROS production by CII in the steady state, although affects on the rates of electron transfer in the thermodynamic cycle including heme *b* (**Figure 6**). The effect of changes in $E_m(b)$ values from −185 mV used in the present model for bovine heart CII (**Supplementary Table S2**) up to +36 mV for *Escherichia coli* (Anderson et al., 2014) on the stationary rates of electron transfer in CII and reduction of heme *b* was studied computationally.

Figure 6A shows that the stationary concentration of reduced heme *b*, b^-, increases slowly up to the small concentration, 12 μM, that is about 5% of the total heme *b* concentration that equal to 235 μM, with an increase in the succinate concentration at $E_m(b) = -185$ mV while increasing $E_m(b)$ up to +36 mV results in almost complete reduction of heme *b*, an increase in b^- concentration up to 230 μM, or 98% of the total heme *b* concentration, at 40 μM succinate concentration. However, despite a strong difference in the reduced steady-state heme *b* concentration at different $E_m(b)$ values the rates of QH_2 production, v_{20}, i.e., the SQR activity, ROS production, and vH_2O_{2tot}, remain unchanged (**Figure 6B**).

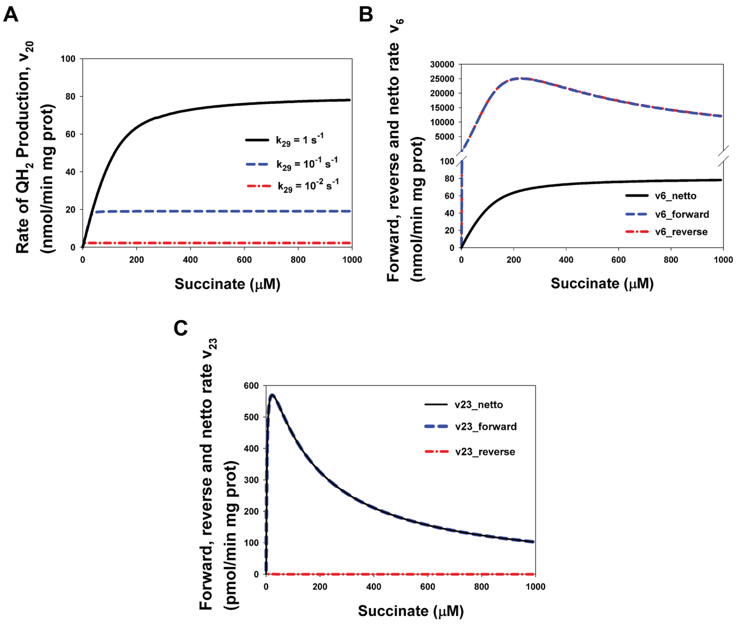


FIGURE 5 | Computer simulated bell-shaped vs. hyperbolic kinetics of ROS production and succinate-Q reductase (SQR) activity of CII. **(A)** The hyperbolic stationary dependence of the QH_2 production rate, v_{20}, on the succinate concentration in the basal state ($k_{29} = 1\ s^{-1}$) and upon intermediate ($k_{29} = 0.1\ s^{-1}$) and strong ($k_{29} = 0.01\ s^{-1}$) inhibition of respiratory Complex III (CIII) that was simulated by decreasing the catalytic constant k_{29} of QH_2 oxidation. **(B)** Computer simulated dependence of stationary rates of unidirectional mainstream electron flows in forward (from $FADH_2$ to oxidized [2Fe-2S] cluster), $v_{6\_forward}$, and reverse (from reduced $[2Fe\text{-}2S]^-$ cluster to $FADH^\bullet$), $v_{6\_reverse}$, direction and netto rate $v_{6\_netto} = v_{6\_forward} - v_{6\_reverse}$ on the succinate concentration. **(C)** Computer simulated dependence of stationary rates of unidirectional bypass electron flows in forward (from $FADH_2$ to O_2), $v_{23\_forward}$, and reverse (from O_2^- to $FADH^\bullet$), $v_{23\_reverse}$, direction and netto rate $v_{23\_netto} = v_{23\_forward} - v_{23\_reverse}$ on the succinate concentration. Values of the rate constant k_{29} shown in **Figure 3A**. Black solid curve corresponds to the $k_{29} = 1\ s^{-1}$, blue dashed curves – $k_{29} = 0.1\ s^{-1}$, and red dash-dot curves – $k_{29} = 0.01\ s^{-1}$. The rest model parameter values are presented in **Supplementary Table S2**. Curves in **Figures 3B,C** are designated as follows: black solid curves correspond to the netto rates v_{6_netto} **(B)** and v_{23_netto} **(C)**, blue dashed curves – unidirectional forward rates $v_{6\_forward}$ **(B)** and $v_{23\_forward}$ **(C)**, and red dash-dot curves – unidirectional reverse rates $v_{6\_reverse}$ **(B)** and $v_{23\_reverse}$ **(C)**. All computer simulations were made at model parameter values presented in **Supplementary Table S2**.

As to the rates of electron transfer related to the thermodynamic cycle [3Fe-4S]↔heme *b*↔Q↔[3Fe-4S], that is reactions (9, 11, and 13) for the first electron and reactions (17, 18, and 21) for the second electron transfer, the computer simulation results presented in **Figures 6C,D** for $E_m(b) = -185$ mV and **Figures 6E,F** for $E_m(b) = +36$ mV show that these rates change very much when changing $E_m(b)$. First of all, it should be pointed that input electron flow entering the thermodynamic cycle from [4Fe-4S] cluster, v_8 for the first electron and v_{16} for the second electron transfer, remain unchanged at different $E_m(b)$ values and equal to the output flow, i.e., the rate of QH_2 production, v_{20}. Other curves in **Figures 6C–F** describe the rates of electron flows inside the thermodynamic cycle. The most interesting of them is v_{11} at $E_m(b) = +36$ mV (**Figure 6E**) that describes the first electron transfer between heme *b* and Q. Computer simulation results predict that the first electron transfer occurs in the reverse direction from semiquinone to heme *b* at the high value of $E_m(b)$ equal to +36 mV. However, this effect is compensated by increasing the first electron flow in forward direction from [3Fe-4S] cluster to Q, v_{13}.

It is necessary to emphasize that it was taken into account "the principle of detail balancing" (Hearon, 1953) for thermodynamic cycles at all computations that requires the product of equilibrium constants along a cycle to be equal to 1. For the thermodynamic cycle [3Fe-4S]↔heme *b*↔Q↔[3Fe-4S] this means the following relations: $K_{eq13} = K_{eq9}{\cdot}K_{eq11}$ and $K_{eq21} = K_{eq17}{\cdot}K_{eq18}$. The values of K_{eq9}, K_{eq11}, and K_{eq17}, K_{eq18} at $E_m(b) = -185$ mV are presented in **Supplementary Table S2**. Increasing $E_m(b)$ from −185 up to +36 mV results in the following changes in these equilibrium constants: $K_{eq9} = K_{eq17} = 5.55{\bullet}10^{-5}$ exp. [(185 + 36) F/RT] = $5.55{\bullet}10^{-5}{\bullet}6{,}905 = 0.38$; $K_{eq11} = 2.72/6905 = 3.94{\bullet}10^{-4}$; and $K_{eq18} = 3.269{\bullet}10^{6}/6905 = 473.4$.

That is, the values of equilibrium constants K_{eq13} and K_{eq21} kept unchanged while K_{eq9}, K_{eq11}, and K_{eq17}, K_{eq18} are varied with increasing $E_m(b)$ value.

These results imply that a high level of ROS production by CII induced by mutations in SDH cyt *b* (Ishii et al., 1998) is not related with changes in $E_m(b)$. It is most likely as was pointed earlier (Tran et al., 2007; Anderson et al., 2014) that heme *b* plays more the structural role stabilizing CII as a heterotetramer than for catalysis in CII as an electron carrier. That is, the heme *b* plays an important role in assembly CII because as was shown in (Yu et al., 1987), *in vivo*, SDH is anchored to the inner membrane with the cytochrome b_{560}.

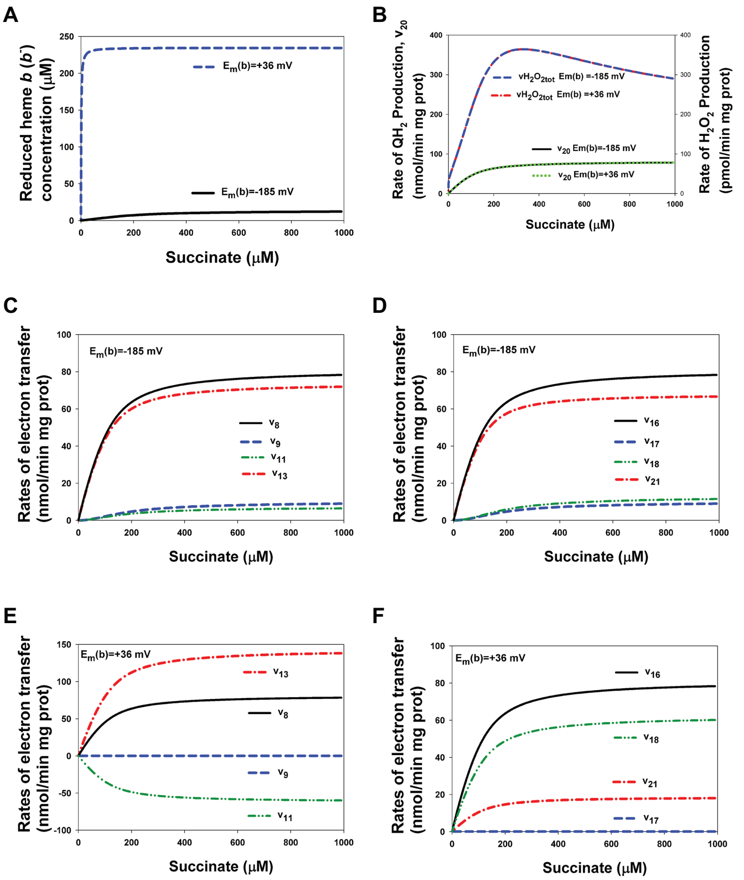


FIGURE 6 | Computer simulated alterations in the reduced heme *b* concentration and rates of electron flows in CII related to changes in the value of heme *b* midpoint potential, $E_m(b)$, from −185 up to +36 mV. **(A)** The dependence of the reduced heme *b*, b^-, concentration on the succinate concentration at two different values of $E_m(b)$, −185 mV (black solid curve) and +36 mV (blue dashed curve). **(B)** The dependence of the rate of QH_2 production, v_{20}, and the total rate of H_2O_2 generation by CII, vH_2O_{2tot}, at two different values of $E_m(b)$, −185 mV (black solid and blue dashed curves, respectively) and + 36 mV (green dot and red dash-dot curves, respectively). **(C–F)** Computer simulated dependence of the rates of electron flows from [4Fe-4S] cluster to ubiquinone at ubiquinone (Q)-binding site on the succinate concentration is shown at two different values of $E_m(b)$, −185 mV **(C,D)** and +36 mV **(E,F)** and separately for the first **(C,E)** and second **(D,F)** electron transfer. All computer simulations were made at model parameter values presented in **Supplementary Table S2**.

Thus, mutations in the cytochrome b large subunit (SDHC), of CII, that induce oxidative stress and lead to apoptosis (see for review Ishii et al., 2007) results in a suppression of the SQR activity of assembled CII due to its disintegration and it is very likely that in this case the dependence of stationary rates of ROS production on the succinate concentration has the same features as upon inhibition of CIII or Q-binding site considered above.

Assembled and Disintegrated CII

CII disintegration resulting in dissociation of the SDHA/SDHB subunits from the membrane-bound SDHC/SDHD complex to the mitochondrial matrix (reviewed in Hwang et al., 2014) and a physical break in electron transfer from succinate to Q at the Q-binding site of CII, i.e., impairment of the SQR activity, causes crucial changes in the kinetics of ROS production by redox centers of these matrix subunits. Comparative analysis of the dependence of the stationary rates of ROS production by different redox centers of CII in assembled and disintegrated states is presented in **Figure 7**. Here, it was considered the case when a part of CII is in the assembled (CIIt = 100 μM) and a part is in the disintegrated state (ABt = CDt = 135 μM). In addition, it should point those stationary rates of ROS production in assembled CII were computed at simulation the basal state of CII without any inhibitors (all parameter values are presented in **Supplementary Table S2**). One can see from **Figure 7** that the kinetics of ROS production by CII changes from sigmoid in assembled state to the high-amplitude bell-shaped kinetics in disintegrated state. These changes qualitatively and quantitatively close to changes in the kinetics of ROS production by assembled CII upon inhibition of CIII or Q-binding site of CII.

One can see from **Figure 7** that the kinetics of ROS production by $FADH_2$ (**Figures 7A,B**) changes from sigmoid in the assembled state (v_{22} and v_{23}) to the high-amplitude

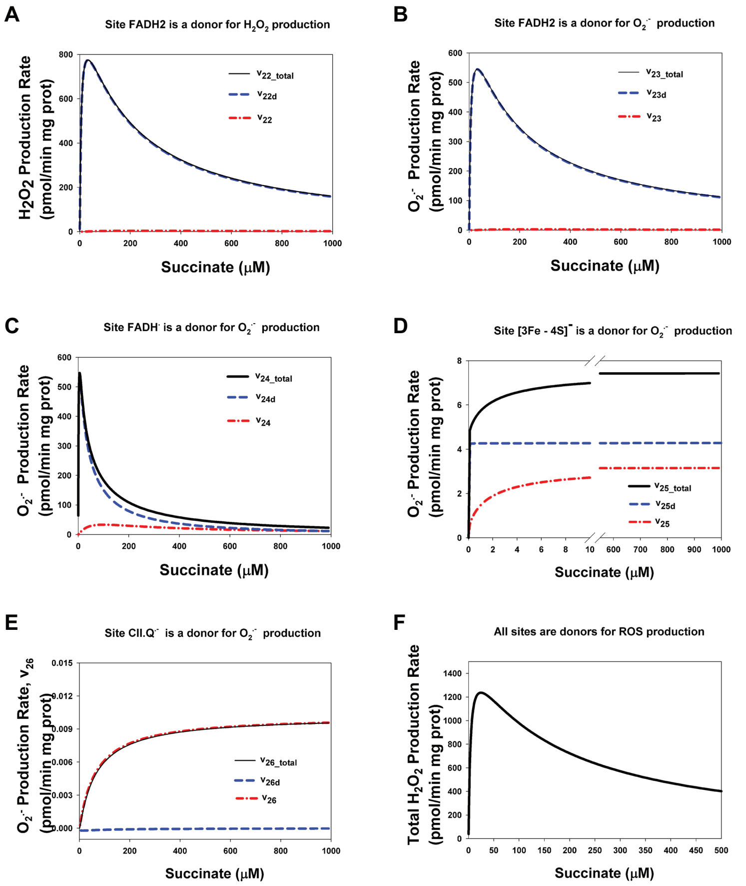


FIGURE 7 | Computer simulated stationary rates of O_2^- and H_2O_2 production by different sites in assembled and disintegrated CII. **(A)** The rate of H_2O_2 production by $FADH_2$ in assembled CII (v_{22}), disintegrated CDHA/CDHB subcomplex of CII (v_{22d}), and the total rate by both these sites (v_{22_total}); **(B–E)** The rates of O_2^- production by the following sites: **(B)** $FADH_2$ (v_{23}), **(C)** $FADH^{\cdot}$ (v_{24}), **(D)** Reduced $[3Fe\text{-}4S]^-$ cluster (v_{25}), and **(E)** Semiquinone anion binding to CII, site $CII.Q^-$ (v_{26}). **(F)** The total rate of ROS production by the all redox centers of assembled and disintegrated CII at different values of the concentration of AA5. The total rate of H_2O_2 production by CII (vH_2O_{2tot}) that was computed as the rate of H_2O_2 release from the mitochondrial matrix to cytosol that equal to the summary rate of H_2O_2 production by $FADH_2$ in assembled, v_{22}, and disintegrated, v_{22d}, states and dismutation of O_2^-, v_{28}, in the matrix at the steady state (see *Explicit functions* in Mathematical model in **Supplementary Data**). All computer simulations were made at model parameter values as for **Figure 2** except AA5 = 0 and the total concentration of CII in the assembled state equal to100 µM and in the disintegrated state ABt = CDt = 135 µM, so the total CII concentration is equal to 235 µM. **(A–E)** Black solid curves correspond to the total ROS production rate for each site in assembled and disintegrated CII, blue dashed curves correspond to the ROS production rate by each site in disintegrated CII, and red dash-dot curves – assembled CII.

bell-shaped kinetics in disintegrated state (v_{22d} and v_{23d}). These changes qualitatively are the same as the changes in the kinetics of ROS production by $FADH_2$ in assembled CII upon inhibition of CIII or Q-binding site of CII (**Figures 3A,B**; **Supplementary Figures S2A,B**). Quantitative difference in these changes is related only to different concentration of $FADH_2$ in the assembled and disintegrated states. Practically, the same changes from close to sigmoid in the assembled state to the high-amplitude bell-shaped kinetics in the disintegrated state occur in the rate of O_2^- production by $FADH^{\bullet}$ (**Figure 7C**). So, the total rate of ROS production by $FADH_2$ ($v_{22\_total} = v_{22} + v_{22d}$) and $FADH^{\bullet}$ ($v_{23\_total} = v_{23} + v_{23d}$) in both the assembled and disintegrated states is high amplitude and bell-shaped due to the main contribution of disintegrated $FADH_2$ and $FADH^{\bullet}$ to the ROS production rate.

The dependence of the rate of O_2^- production by reduced $[3Fe\text{-}4S]^-$ cluster on the succinate concentration keeps hyperbolic shape under any condition (**Figure 5D**) with a decrease in the Michaelis constant in the disintegrated state like in assembled state upon inhibition of the Q-site by AA5 (**Figure 3D**). And, as expected, the total rate of O_2^- production by semiquinone, $CII.Q^-$, at the Q-binding site also keeps hyperbolic shape in the dependence on the succinate concentration (**Figure 7E**) and includes ROS production in the assembled state only.

The total rate of ROS production by the all sites of assembled and disintegrated CII at different values of the rate constant k_{10} that simulate the effect of AA5 is presented in **Figure 7F**. The total rate of H_2O_2 production by CII (vH_2O_{2tot}) that was computed as the rate of H_2O_2 release from the mitochondrial matrix to cytosol is indistinguishable by ROS generation source, assembled, or disintegrated CII, so it is equal to the summary rate of H_2O_2 production by $FADH_2$ in assembled, v_{22}, and disintegrated, v_{22d}, states and dismutation of O_2^-, v_{28}, in the matrix in the steady state (see *Explicit functions* in Mathematical model in the section **Supplementary Data**). As expected, the maximal rate of ROS production by CII occurs under both conditions, disintegration of CII and upon inhibition of the Q-binding site of assembled CII.

The dependence of stationary rates of ROS production by CII in completely disintegrated state on the succinate concentration is presented in **Supplementary Figure S4**. Completely disintegrated CII means that all CDHA/CDHB subcomplexes of CII that produce ROS are dissociated from the membrane and located only in the matrix with the concentration of 235 μM. All of dependencies of the ROS production rate by different sites of completely disintegrated CII on the succinate concentration match those in the assembled CII upon a strong inhibition of CIII and the Q-binding site.

Thus, these computer simulation results predict that CII disintegration results in the same changes in the kinetics of ROS production by CII as a suppression of SQR activity of assembled CII upon inhibition of the Q-binding site and/or CIII as was pointed earlier in (Siebels and Dröse, 2013) and induces the high-amplitude bell-shaped dependence of the rate of ROS production by $FADH_2$ in the unoccupied dicarboxylate state and $FADH^•$ with a shift of the maximal rate to small subsaturated concentration of succinate.

CONCLUSION

A computational, mechanistic model of electron transfer and the formation of superoxide (O_2^-) and hydrogen peroxide (H_2O_2) in CII in the assembled and disintegrated states was developed in the present study to facilitate quantitative analysis of mitochondrial ROS production. The model was calibrated by fitting the computer simulated results to experimental data obtained on SMP prepared from bovine heart mitochondria upon inhibition of Q-binding site by AA5 (Siebels and Dröse, 2013) and from rat heart mitochondria upon inhibition of Complex III by myxothiazol (Grivennikova et al., 2017).

The present computational modeling study predicts that a suppression of the SQR activity of CII resulting from inhibition of CIII or Q-binding site of CII as was pointed earlier in (Siebels and Dröse, 2013) as well as CII disintegration reviewed in (Hwang et al., 2014) causes transition in the succinate-dependence of ROS production from small-amplitude sigmoid (hyperbolic) determined by Q-binding site or/and [3Fe-4S] cluster to the high-amplitude bell-shaped kinetics with a shift to small subsaturated concentration of succinate determined by $FADH_2$ in the unoccupied dicarboxylate state and $FADH^•$.

Computer simulation results confirm previous hypothesis (Quinlan et al., 2012; Siebels and Dröse, 2013) that the main contribution to the total rate of ROS production by CII upon inhibition of the Q-binding site or/and CIII give unoccupied $FADH_2$ and $FADH^•$.

The dependence of the rates of ROS production by disintegrated matrix SDHA/SDHB subcomplexes on the succinate concentration matches qualitatively and quantitatively to those in the assembled state upon inhibition of the CII Q-binding site or/and CIII.

It is very likely that semiquinone at the Q-binding site and [3Fe-4S] cluster give a small contribution to ROS production in both basal and inhibited state. Presented modeling results show that the dependence of the rates of ROS production by these redox centers on the succinate concentration keeps hyperbolic shape with very small maximal rate due to the small values of the catalytic constants of ROS formation in any state of CII.

Computational modeling analysis of the model included the thermodynamic cycle [3Fe-4S]↔heme *b*↔Q↔[3Fe-4S] shows that the value of midpoint potential of the heme *b*, $E_m(b)$, does not affect on the rates of QH_2 and ROS production by CII in the steady state, although affects on the heme *b* reduction and the rates of electron transfer in the thermodynamic cycle including heme *b*. This result confirms theoretically suggestions pointed earlier (Tran et al., 2007; Anderson et al., 2014) that heme *b* plays more the structural role stabilizing CII as a heterotetramer than for catalysis in CII as an electron carrier.

Thus, the results of this work allow us to evaluate the catalytic constants of ROS formation by each of the redox centers of CII and predict their contribution to the overall generation of ROS by CII in various (basal or inhibited as well as assembled or disintegrated) states of CII. We hope that this will help resolve the many years of debate about which of the CII redox centers are actually involved in the ROS formation. These theoretical predictions are particularly valuable now that experimental methods for directly measuring these catalytic constants have not yet been developed.

Theoretical results on ROS production by disintegrated matrix SDHA/SDHB subcomplexes with bell-shaped dependence on succinate concentration are extremely useful for understanding autocatalytic MPT due to efflux of succinate from mitochondria through the open MPT pore found recently in cardiac mitochondria (Korge et al., 2017). In this case, activation of ROS production by CII at low succinate concentration results in activation of MPT with following activation of apoptosis or necrosis/necroptosis.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

NM and JH conceived and supervised the study. NM and LM performed computations. NM and JH wrote this report. All authors contributed to the article and approved the submitted version.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The work was carried out with budget funding under the State assignment of the Institute of Theoretical and Experimental Biophysics of the Russian Academy of Sciences (ITEB RAS) No. 075-00845-20-01. We are grateful to Administration of ITEB RAS for covering open access publication fee.

ACKNOWLEDGMENTS

We thank Dr. M.H. Galimova for helpful discussion of the computational results.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.557721/full#supplementary-material

REFERENCES

Anderson, R. F., Shinde, S. S., Hille, R., Rothery, R. A., Weiner, J. H., Rajagukguk, S., et al. (2014). Electron transfer pathways in the heme and quinine-binding domain of complex II. *Biochemistry* 53, 1637–1646. doi: 10.1021/bi401630m

Grivennikova, V. G., Kozlovsky, V. S., and Vinogradov, A. D. (2017). Respiratory complex II: ROS production and the kinetics of ubiquinone reduction. *Biochim. Biophys. Acta* 1858, 109–117. doi: 10.1016/j.bbabio.2016.10.008

Hearon, J. Z. (1953). The kinetics of linear systems with special reference to periodic reactions. *Bull. Math. Biophys.* 15, 121–141. doi: 10.1007/BF02476377

Hwang, M. -S., Rohlena, J., Dong, L. -F., Neuzil, J., and Grimm, S. (2014). Powerhouse down: complex II dissociation in the respiratory chain. *Mitochondrion* 19, 20–28. doi: 10.1016/j.mito.2014.06.001

Ishii, N., Fuji, M., Hartman, P. S., Tsuda, M., Yasuda, K., Senoo-Matsuda, N., et al. (1998). A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. *Nature* 394, 694–697. doi: 10.1038/29331

Ishii, N., Ishii, T., and Hartman, P. S. (2007). The role of the electron transport SDHC gene on lifespan and cancer. *Mitochondrion* 7, 24–28. doi: 10.1016/j.mito.2006.11.012

Korge, P., John, S. A., Calmettes, G., and Weiss, J. N. (2017). Reactive oxygen species production induced by pore opening in cardiac mitochondria: the role of complex II. *J. Biol. Chem.* 292, 9896–9905. doi: 10.1074/jbc.M116.768325

Lemarie, A., Huc, L., Pazarentzos, E., Mahul-Mellier, A. -L., and Grimm, S. (2011). Specific disintegration of complex II succinate: ubiquinone oxidoreductase links pH changes to oxidative stress for apoptosis induction. *Cell Death Differ.* 18, 338–349. doi: 10.1038/cdd.2010.93

Markevich, N. I., Galimova, M. H., and Markevich, L. N. (2019). Mathematical model of electron transfer and formation of reactive oxygen species in mitochondrial complex II. *Biochem (Mosc). Suppl. Ser A. Membr. Cell Biol.* 13, 341–351. doi: 10.1134/S199074781904007X

McLennan, H. R., and Degli Esposti, M. (2000). The contribution of mitochondrial respiratory complexes to the production of reactive oxygen species. *J. Bioenerg. Biomembr.* 32, 153–162. doi: 10.1023/a:1005507913372

Moosavi, B., Berry, E. A., Zhu, X. L., Yang, W. C., and Yang, G. F. (2019). The assembly of succinate dehydrogenase: a key enzyme in bioenergetics. *Cell. Mol. Life Sci.* 76, 4023–4042. doi: 10.1007/s00018-019-03200-7

Quinlan, C. L., Orr, A. L., Perevoshchikova, I. V., Treberg, J. R., Ackrell, B. A., and Brand, M. D. (2012). Mitochondrial complex II can generate reactive oxygen species at high rates in both the forward and reverse reaction. *J. Biol. Chem.* 287, 27255–27264. doi: 10.1074/jbc.M112.374629

Ristow, M., and Schmeisser, K. (2014). Mitohormesis: promoting health and lifespan by increased levels of reactive oxygen species (ROS). *Dose-Response* 12, 288–341. doi: 10.2203/dose-response.13-035.Ristow

Schwall, C. T., Greenwood, V. L., and Alder, N. N. (2012). The stability and activity of respiratory complex II is cardiolipin-dependent. *Biochim. Biophys. Acta* 1817, 1588–1596. doi: 10.1016/j.bbabio.2012.04.015

Siebels, I., and Dröse, S. (2013). Q-site inhibitor induced ROS production of mitochondrial complex II is attenuated by TCA cycle dicarboxylates. *Biochim. Biophys. Acta* 1827, 1156–1164. doi: 10.1016/j.bbabio.2013.06.005

Tran, Q. M., Rothery, R. A., Maklashina, E., Cecchini, G., and Weiner, J. H. (2007). Ischerichia coli succinate dehydrogenase variant lacking the heme b. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18007–18012. doi: 10.1073/pnas.0707732104

Turrens, J. F. (2003). Mitochondrial formation of reactive oxygen species. *J. Physiol.* 552, 335–344. doi: 10.1113/jphysiol.2003.049478

Yu, L., Xu, J. -X., Haley, P. E., and Yu, C. -A. (1987). Properties of bovine heart mitochondrial cytochrome b560. *J. Biol. Chem.* 262, 1137–1143.

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fphys.2020.591073

A Computational Model for the Cold Response Pathway in Plants

Ruqiang Zhang[1], Didier Gonze[2], Xilin Hou[1], Xiong You[3]* and Albert Goldbeter[2]*

[1]College of Horticulture, Nanjing Agricultural University, Nanjing, China, [2]Unité de Chronobiologie Théorique, Faculté des Sciences, Université Libre de Bruxelles (ULB), Brussels, Belgium, [3]College of Sciences, Nanjing Agricultural University, Nanjing, China

Understanding the mechanism by which plants respond to cold stress and strengthen their tolerance to low temperatures is an important and challenging task in plant sciences. Experiments have established that the first step in the perception and transduction of the cold stress signal consists of a transient influx of Ca^{2+}. This Ca^{2+} influx triggers the activation of a cascade of phosphorylation-dephosphorylation reactions that eventually affects the expression of C-repeat-binding factors (CBFs, notably CBF3), which were shown in many plants to control resistance to cold stress by regulating the expression of cold-regulated (COR) genes. Based on experimental observations mostly made on *Arabidopsis thaliana*, we build a computational model for the cold response pathway in plants, from the transduction of the cold signal *via* the transient influx of Ca^{2+} to the activation of the phosphorylation cascade leading to CBF3 expression. We explore the dynamics of this regulatory network by means of numerical simulations and compare the results with experimental observations on the dynamics of the cold response, both for the wild type and for mutants. The simulations show how, in response to cold stress, a brief Ca^{2+} influx, which is over in minutes, is transduced along the successive steps of the network to trigger the expression of cold response genes such as CBF3 within hours. Sometimes, instead of a single Ca^{2+} spike the decrease in temperature brings about a train of high-frequency Ca^{2+} oscillations. The model is applied to both types of Ca^{2+} signaling. We determine the dynamics of the network in response to a series of identical cold stresses, to account for the observation of desensitization and resensitization. The analysis of the model predicts the possibility of an oscillatory expression of CBF3 originating from the negative feedback exerted by ZAT12, a factor itself controlled by CBF3. Finally, we extend the model to incorporate the circadian control of CBF3 expression, to account for the gating of the response to cold stress by the plant circadian clock.

Keywords: cold response pathway, cold tolerance, C-repeat-binding factor, computational model, systems biology, plant cold acclimation

Edited by:
Jianhua Xing,
University of Pittsburgh,
United States

Reviewed by:
Xiaojun Tian,
Arizona State University,
United States
Kevin P. Vincent,
University of California, San Diego,
United States

***Correspondence:**
Albert Goldbeter
agoldbet@ulb.ac.be
Xiong You
youx@njau.edu.cn

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: *03 August 2020*
Accepted: *16 October 2020*
Published: *05 November 2020*

Citation:
Zhang R, Gonze D, Hou X, You X and Goldbeter A (2020) A Computational Model for the Cold Response Pathway in Plants.
Front. Physiol. 11:591073.
doi: 10.3389/fphys.2020.591073

INTRODUCTION

Low temperature has adverse effects on the survival, growth, and development of plants (Chew and Halliday, 2010). In order to survive exposition to low temperature, plants have evolved sophisticated mechanisms to sense the seasonal, daily, and rapid fluctuations in temperature, and to adjust their physiology appropriately (Kaplan et al., 2004; Chinnusamy et al., 2006).

In many temperate plants, cold acclimation increases freezing tolerance after exposure to nonfreezing low temperatures (Thomashow, 1999; Chinnusamy et al., 2007; Ding et al., 2019). It is essential to unravel the molecular mechanism of cold sensing and cold stress signal transduction to avoid damage in plants caused by low temperatures. Computational models closely related to experimental observations can prove helpful to clarify the dynamics of signal transduction pathways at the cellular level. The goal of this paper will be to present such a computational model for the cold response pathway in plants.

Experimental studies in *Arabidopsis* have led to the identification of key factors in the transcriptional network of the cold acclimation pathway, where three C-repeat (CRT)-binding factors (CBFs), the most commonly studied being CBF3, also known as dehydration-responsive element (DRE)-binding proteins (DREBs), play vital roles in cold acclimation (Stockinger et al., 1997; Jaglo-Ottosen et al., 1998; Liu et al., 1998; Zhao et al., 2016). These transcription factors are induced by cold stress and bind to CRT/DRE DNA regulatory elements in the promoters of a large subset of cold-regulated (COR) genes. The expression of these *COR* genes renders plants able to tolerate freezing stress through a variety of cellular regulatory mechanisms (Thomashow, 1999; Gilmour et al., 2004; Hannah et al., 2006; Lissarre et al., 2010; Barah et al., 2013; Liu et al., 2019).

Homologues of *CBF* genes have been found in different plant species, such as the Chinese cabbage *Brassica campestris* (Wang et al., 2014), *Brassica napus* (Jaglo et al., 2001), barley (Marozsán-Tóth et al., 2015), tomato (Zhang et al., 2004), and rice (Dubouzet et al., 2003). Their functions are similar to those of the *Arabidopsis* genes, and they are also induced in response to low temperature. Given that the components of the cold response pathway are highly conserved in many plant species and that the patterns of expression of the *CBF* and *COR* genes correspond to those observed in *Arabidopsis*, the model that we propose for the cold response pathway pertains not only to *Arabidopsis thaliana* but also to other plants.

The putative temperature sensing mechanisms are able not only to respond to temperature changes but also to activate downstream response pathways (Ruelland and Zachowski, 2010). Calcium (Ca^{2+}), a widely studied plant second messenger, is involved in nearly every aspect of cell physiology and development (Kudla et al., 2018). Cytosolic Ca^{2+} alterations and oscillations in plant cells, induced by a variety of environmental stimuli, are an integral component of cell signaling, and the frequency, amplitude, and spatial localization of Ca^{2+} signals control the efficiency and specificity of cellular responses (Allen et al., 2001). In plants, guard cells integrate environmental and endogenous signals to regulate the aperture of stomatal pores. Cytosolic Ca^{2+} oscillations are essential for stomatal closure, which follows cytosolic Ca^{2+} elevation in guard cell (Allen et al., 2001). Relationships between temperature sensing, specifically membrane fluidity, and Ca^{2+} signaling have been reported (McClung and Davis, 2010). The cold response in plants that follows a cold stress is strongly dependent on a fast and transient cytosolic Ca^{2+} elevation (Knight et al., 1996; Örvar et al., 2000; Cook et al., 2004).

The molecular mechanisms involved in the cold signal transduction, which include the perception of the signal, the cascade of post-translational modifications triggered by the signal, and the transcriptional regulatory network underlying cold acclimation, have all been investigated at the molecular level (Hua, 2009; Shi et al., 2018). To obtain additional insights into the dynamics of the cold response, it is useful to develop a detailed computational model that incorporates the sequence of biochemical processes involved in the signaling pathway. Such a model should eventually prove useful to interpret quantitatively the available experimental findings and to guide further experimental studies. Based on the experimental observations that revealed the structure of the pathway, we build a detailed computational model with the aim of providing a framework for a quantitative description of the regulatory network that controls the plant response to cold stress. The lack of detailed quantitative experimental observations poses a challenge in regard to parameter optimization. Most observations pertain to the time course of the expression of the gene coding for CBF3 in response to cold stress, while the time course of expression of a few other genes of the pathway has also been measured. We compare the predictions of the model with the available experimental data both in wild type and mutants.

In "The Cold Response Pathway: Brief Overview of Experimental Aspects" section, we summarize the processes underlying the response of plants to cold stress. The initial signal triggered by the cold stress takes the form of a brief influx of Ca^{2+}, which rises in seconds and is over in a few minutes. This signal activates a cascade of enzymatic reactions that leads, within a few hours, to the expression of *CBF3* mRNA, which represents the major output of the pathway that is measured in experiments. Based on the regulations determined experimentally, we present, in "Modeling the Plant Response to Cold Stress" section, a computational model for the plant cold response pathway. In "Dynamics of the Response to Cold Stress: Model Predictions" section, by means of numerical simulations, we determine the dynamical evolution of the components of the network following a cold stress. The numerical results of the computational model are compared with the available experimental observations in both wild type and mutants. Under certain conditions, the model predicts the possible occurrence of an oscillatory expression of *CBF3* in response to cold stress. This potential periodic phenomenon originates from the negative feedback exerted on *CBF3* expression by ZAT12, a factor itself controlled by CBF3. An additional level of complexity arises from the control exerted by the plant circadian clock on the cold response pathway. The *CBF3* expression under natural conditions of a light-dark (LD) cycle is indeed circadian and peaks at about ZT8, i.e., 8 h after light onset, just after the early morning peak of the proteins CIRCADIAN CLOCK-ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) involved in the circadian clock mechanism (Dong et al., 2011). At the end of "Dynamics of the Response to Cold Stress: Model Predictions" section, we will indicate how the model for the cold response pathway can be extended to incorporate its modulation by the circadian clock, which is responsible for the gating of the plant response to cold stress. We discuss the results in "Discussion" section where we also allude to vernalization, another plant response to cold, which, in contrast to the response leading to freezing tolerance, occurs on a much longer time scale.

THE COLD RESPONSE PATHWAY: BRIEF OVERVIEW OF EXPERIMENTAL ASPECTS

The plant cold signaling pathway has been extensively studied in the past two decades (Chinnusamy and Zhu, 2002; Hannah et al., 2005; Thomashow, 2010; Shi et al., 2018). The salient features of the cold response pathway used in building the computational model are presented schematically in **Figure 1**. Below, we present in turn the messenger molecules, the protein kinases and phosphatases, and the transcription factors involved in this pathway.

Cold Stress Sensing

Under cold shock, when the temperature drops from normal to chilling, the cytosolic concentration of Ca^{2+} ($[Ca^{2+}]$) in plant cells shows a sharp rise before returning to a basal level. The steady-state cytoplasmic calcium concentration in plants is in

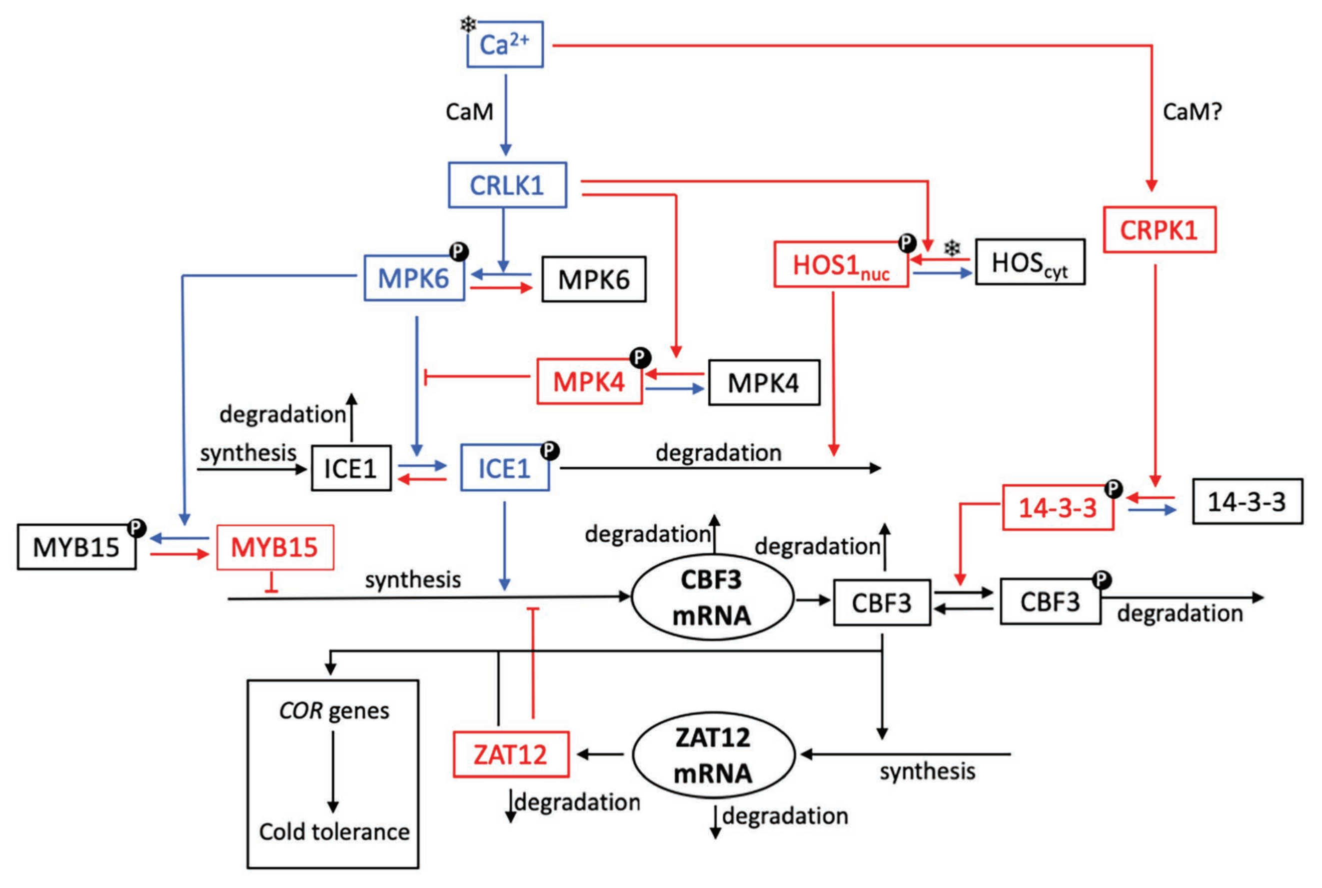


FIGURE 1 | Schematic model for the cold response pathway in plants. As explained in further detail in "The Cold Response Pathway: Brief Overview of Experimental Aspects" section, the plasma membrane is thought to be the primary target of cold sensing and the starting point for transmission of the cold signal into the nucleus. The cold stress (symbolized by a snowflake) initiates a transient increase in Ca^{2+}, which takes the form of a single Ca^{2+} pulse or of a train of high-frequency Ca^{2+} oscillations. The cold response pathway reacts to the cold stress by the expression of C-repeat-binding factors (CBFs, notably CBF3) that control the resistance of plants to cold stress through the expression of cold-regulated (COR) genes. For the sake of clarity, not all reaction steps are indicated in the scheme. Steps colored in blue and red denote reactions or interactions which contribute to enhance or inhibit CBF3 expression, respectively. Upon activation by calmodulin (CaM), plasma membrane Ca^{2+}-regulated kinases (CRLK1/2) positively regulate cold-induced gene expression, through the MPK6 and MPK4 pathway. The cold signal activates the MPK4-MPK6 cascade to regulate freezing tolerance, through phosphorylation of ICE1. A receptor-like cytoplasmic kinase cold-responsive protein kinase 1 (CRPK1) is activated by low temperature stress and phosphorylates 14-3-3 proteins; in the absence of direct evidence (hence the question mark), we assume that this activation, as for CRLK1, is mediated by CaM (see "CBF Signaling" section). To keep the scheme simple, we do not show the active forms CRLK1a and CRPK1a resulting from the bimolecular activation of CRLK1 and CRPK1 by CaM. The key components of CBF-dependent signaling, ICE1 and CBF, are modulated through post-translational modifications. Phosphorylated MPK6 mediates the phosphorylation of ICE1, which leads to increased ICE1 degradation *via* ubiquitination; the latter process is enhanced by the high expression of the osmotically responsive gene1 (HOS1). The 14-3-3 protein kinase phosphorylates CBF3 and, thereby, facilitates the ubiquitin-mediated degradation of CBF3. Cold stress induces the accumulation of CBF proteins, which elicit the expression of ZAT12. The latter factor exerts a negative feedback on the transcription of *CBF3* mRNA.

the range of 80~250 nM (Knight et al., 1996). The [Ca^{2+}] changes that occur during cooling are a complex function of the rate of cooling, the duration of cooling, and the magnitude of the temperature drop (Knight et al., 1996; Plieth et al., 1999). The elevation in cytosolic Ca^{2+} represents the primary signal, which is transmitted through Ca^{2+}-regulated proteins, and in turn affects the phosphorylation of various proteins. The major Ca^{2+} sensor in plants in the cell plasma membrane is calmodulin (CaM), which is activated through the binding of Ca^{2+} in a cooperative manner, and ultimately regulates, *via* a signaling cascade, the expression of target genes (Zielinski, 1998). The decoding of different Ca^{2+} signatures causes the changes in gene expression that lead to the appropriate physiological responses (Knight et al., 1996; Plieth, 1999; Plieth et al., 1999).

MAPKs Signaling

Cold stress-induced Ca^{2+} transient changes resulting in CaM activation can be decoded by different pathways. When the plant senses low temperature by an uncharacterized Ca^{2+}/calmodulin-regulated receptor-like kinase (CRLK), the autophosphorylation of this CRLK —particularly CRLK1—is activated. The activation of CRLK1 leads to the phosphorylation of downstream targets such as the mitogen-activated protein kinases 6 (MPK6) and 4 (MPK4), which elicits the cold response (Yang et al., 2010; Furuya et al., 2013; Zhao et al., 2017). The MEKK1-MKK2-MPK4 cascade constitutively reduces the protein levels and kinase activities of MPK3 and MPK6 (Teige et al., 2004; Kong et al., 2012). The latter two kinases phosphorylate inducer of CBF expression 1 (ICE1), a transcription factor that regulates the expression of *CBF* genes, while the phosphorylation of ICE1 promotes its degradation (Zhao et al., 2017). The MEKK1-MKK2-MPK4 pathway constitutively suppresses MPK3 and MPK6 activities and promotes the cold response (Zhao et al., 2017).

CBF Signaling

The expression of *CBF* genes is required for freezing tolerance in *A. thaliana* (Chinnusamy et al., 2003; Agarwal et al., 2006; Doherty et al., 2009). The expression of *CBF* genes is positively regulated by ICE1 and negatively regulated by MYB15. These transcription factors directly interact with specific elements in the *CBF* promoters. MAPK/MPK cascades function upstream to regulate *CBFs*. The downstream ZAT12 regulon appears to be involved in negative regulation of the CBF cold response pathway (Vogel et al., 2005). The *CBF3* transcript has a half-life of only 7.5 min at warm temperatures, a value that is among the shortest described for plant genes (Zarka et al., 2003).

The factor ICE1 is a major regulator that controls *CBF* expression and many other cold-responsive regulons (Chinnusamy et al., 2003). Indeed, ICE1 binds to *MYC* regulatory elements in the *CBF3* promoter and, thereby, induces the expression of *CBF3* during cold acclimation (Tang et al., 2020). The protein HOS1 negatively regulates ICE1 at low temperature by inducing its ubiquitination-mediated degradation (Dong et al., 2006). At normal growth temperature, HOS1 resides in the cytoplasm, while the nuclear localization of HOS1 is enhanced by cold stress (Ishitani et al., 1998; Lee et al., 2001). ICE1 degradation through the 26S proteasome pathway is induced by HOS1. MPK3/MPK6 phosphorylate and destabilize ICE1 and, thereby, negatively regulate *CBF* expression and the freezing tolerance in plants (Miura et al., 2007; Li et al., 2017; Zhao et al., 2017; Tang et al., 2020).

MYB15, a transcriptional repressor of cold signaling, is phosphorylated by phosphorylated MPK6 (Agarwal et al., 2006). Under normal temperatures, *CBF* transcription is repressed by the DNA-binding of unphosphorylated MYB15 to MYB recognition elements in their promoters in the absence of ICE1 activity. In response to cold stress, MYB15 is phosphorylated by cold-activated MPK6. The phosphorylated MYB15 dissociates from *CBF* promoters because of its reduced DNA-binding affinity (Kim et al., 2017).

The stability of the CBF3 protein is controlled by several processes. The kinase cold-responsive protein kinase 1 (CRPK1) and the 14-3-3 protein kinases play negative roles to prevent excessive cold responses (Liu et al., 2017). Cold-induced CBF protein accumulation is compromised by the plasma membrane CRPK1-mediated phosphorylation of 14-3-3 proteins, indicating a negative feedback function in cold signal transduction from the plasma membrane. Cold stress activates, possibly through phosphorylation by a yet unknown receptor kinase, the plasma membrane-localized protein kinase CRPK1 (Liu et al., 2017; Ding et al., 2019), which phosphorylates 14-3-3 proteins in the cytoplasm, thereby triggering 14-3-3 proteins to translocate into the nucleus. We will assume that, as for CRLK1, the activating effect of cold stress on CRPK1 is mediated by CaM. In the nucleus, phosphorylated 14-3-3 proteins promote the 26S proteasome-mediated degradation of CBF proteins, thus attenuating the CBF response. As overexpression of *CBFs* has a strong negative impact on plant growth (Achard et al., 2008), prolonged cold stress would lead to activation of a negative loop *via* CRPK1 activation. In this case, CRPK1 would ensure proper adjustment of the intensity and/or duration of the cold stress response that would need to match the intensity and/or duration of the initial cold stimulus.

MODELING THE PLANT RESPONSE TO COLD STRESS

Our goal is to develop a detailed computational model for the plant response to cold stress involving the succession of biochemical events, from the initial Ca^{2+} influx to the rise in CBF3 that controls the expression of genes which confer tolerance to low temperatures. The model based on the experimental observations summarized above in "The Cold Response Pathway: Brief Overview of Experimental Aspects" section (where relevant references are indicated) is represented schematically in **Figure 1**. Arrows in blue and red denote, respectively, reactions that contribute either to enhance or to inhibit the expression of CBF3 in response to cold stress.

Starting from the top left, we can follow the reactions that are successively triggered by a cold stress. First, the decrease in temperature elicits a brief transient influx in Ca^{2+}. This initial signal leads to the activation of CaM upon cooperative binding of Ca^{2+} (with a Hill coefficient of 4) to the protein. The activation of CaM brings about the activation of two protein kinases, CRLK1 (left branch) and CRPK1 (right branch). We first describe the series of reactions initiated by the activation of CRLK1. We will return afterwards to the right branch initiated by CRPK1, for which available experimental observations are not as detailed as for CRLK1.

The kinetic equations (4)–(17) governing the time evolution of the 14 state variables of the model are listed in **Supplementary Material**, Section 2. The list of these variables, together with their definition, are given in **Supplementary Table 1**, while parameter definitions and numerical values used in the numerical simulations are listed in **Supplementary Table 2**.

Influx of Ca^{2+} and Activation of Calmodulin After a Cold Shock

Cold stress increases the level of cytosolic Ca^{2+} in plant cells. The rise in $[Ca^{2+}]$ takes the form of a single brief spike, or of a train of spikes. A discrete quantitative description of $[Ca^{2+}]$ dynamics for the generation of a single spike after a drop in temperature was proposed by Plieth (1999). Because we focus on the dynamics of the cold response pathway, we resort to a simpler, more straightforward approach by using an instantaneous increase followed by an exponential decrease in cytosolic Ca^{2+}. Parameters were selected so as to produce a single peak in $[Ca^{2+}]$ of appropriate magnitude and duration (see **Supplementary Material**, Section 1.1).

Sometimes the response to a cold shock takes the form of a train of Ca^{2+} oscillations (Allen et al., 2000, 2001). To model the effect of a train of intracellular Ca^{2+} spikes, as outlined in **Supplementary Material**, Section 1.2, we resort to a two-variable model previously proposed for signal-induced Ca^{2+} oscillations (Goldbeter et al., 1990; Dupont et al., 1991). This widely used, generic two-variable model shows that repetitive Ca^{2+} spikes may be triggered by external stimulation, when taking into account the self-amplified release of Ca^{2+} from intracellular stores, a process known as Ca^{2+}-induced Ca^{2+} release.

The CRLK1 Branch: Promotion (and Also Inhibition) of CBF3 Production

Once activated, the protein kinase CRLK1 activates through phosphorylation the MAP kinase MPK6. As for all phosphorylation reactions considered in the model, this process is reversed by a protein phosphatase. Activated MPK6 phosphorylates and, thereby, activates the protein ICE1, which plays the major role in inducing the expression of the gene coding for CBF3. MPK6 also enhances *CBF3* expression by inactivating through phosphorylation the factor MYB15, which is a repressor of the expression of CBF3. Moreover, CRLK1 also contributes to decrease CBF3 production by activating, through phosphorylation, the MAP kinase MPK4, which inhibits the activation of ICE1 by MPK6.

An additional role of CRLK1, less documented, however, raises the possibility that this kinase, once activated by CaM, may also contribute to decrease the CBF3 response and, thus, the tolerance to low temperatures. This putative action involves the cold-stimulated entry into the nucleus and concomitant activation of the factor HOS1, which mediates the ubiquitination of ICE1 and, thereby, enhances its degradation (Dong et al., 2006).

The CRPK1 Branch: Inhibition of CBF3

Most sensory systems display adaptation to constant stimuli (Koshland et al., 1982). After an initial response to a stimulus, the response decreases in time owing to the initiation of a mechanism that limits or inhibits the initial response. Such a phenomenon is also observed for the cold response pathway. We have already mentioned the inhibition exerted by MPK4 on the activation of ICE1 by MPK6: following its activation by CaM after the cold shock, CRLK1 indeed activates both the CBF3-enhancing MPK6 and the CBF3-inhibiting MPK4. The right branch in the scheme in **Figure 1** provides an additional mechanism for the attenuation of the CBF3 response. Indeed, the Ca^{2+} influx also elicits the activation of the kinase CRPK1, likely *via* CaM. The active kinase CRPK1 phosphorylates and, thereby, activates the kinase 14-3-3 which, in turn, phosphorylates the CBF3 protein and marks it for degradation.

Negative Feedback of ZAT12 on CBF3 Expression

The last step in the cold response pathway consists in the induction by CBF3 of *COR* genes responsible for cold tolerance. The activating effect is mediated by the transcription factor ZAT12, which is synthesized in response to a rise in CBF3. However, ZAT12 additionally acts as a repressor of the expression of the gene coding for CBF3. This regulation creates a negative feedback loop between CBF3 and ZAT12 in the lower part of the pathway (Vogel et al., 2005).

DYNAMICS OF THE RESPONSE TO COLD STRESS: MODEL PREDICTIONS

Dynamics of the Pathway at 20°C

We first examine the behavior of the model at a constant temperature of 20°C. At such temperature, the concentration of Ca^{2+} does not change much, so that we can treat it as constant. When the cold response pathway operates under such constant temperature, for given values of the parameters, the system converges to a unique, stable steady state. By setting the Ca^{2+} concentration to a fixed value, we can determine this steady state numerically. Linear stability analysis performed by means of the XPP program developed by Ermentrout (2002; see http://www.math.pitt.edu/~bard/xpp/xpp.html) indicates that this steady state is stable for the parameter values listed in **Supplementary Table 2**, which correspond to a temperature of 20°C. The Ca^{2+} steady-state level is relatively low at this temperature, so that CRLK1 is predominantly in its inactive form while MPK6, MPK4, ICE1, and MYB15 are mostly unphosphorylated. The high level of active, unphosphorylated MYB15 prevents the transcription of *CBF3* so that before the cold stress, CBF3 remains at a low stable steady-state level.

Dynamics Under a Single Cold Shock: Response to a Single Ca^{2+} Pulse

In the model, we consider the level of *CBF3* mRNA expression as the output measuring the response to cold stress. The model allows us to follow in a detailed manner how all the elements of the cold response pathway, i.e., the state variables of the model, change in time after a drop in temperature, from the initial Ca^{2+} influx to the expression of CBF3 and ZAT12. The advantage of the model is that it permits us to follow the evolution of every state variable, as shown in **Figures 2, 3**, in contrast to experimental studies in which, so far, the time course has only been determined for *CBF3* mRNA and *ZAT12* mRNA. The accumulation and subsequence disappearance of these two mRNA species have been measured over the hours that follow the drop in temperature (Fowler and Thomashow, 2002; Chinnusamy et al., 2003).

In **Figure 2**, we show the time course predicted by the model, after the Ca^{2+} signal (A) triggered by the cold stress, for activated CRLK1 (B), phosphorylated MPK6 (MPK6P; C), phosphorylated ICE1 (ICE1P; D), MYB15 (E), phosphorylated MPK4 (MPK4P; F), nuclear HOS1 (HOS1n; G), *CBF3* mRNA (H), and *ZAT12* mRNA (I). In **Figures 2H,I**, we compare the time courses predicted for *CBF3* mRNA and *ZAT12* mRNA (solid curves) with experimental data (red dots for *CBF3* mRNA and blue dots for *ZAT12* mRNA) available for these state variables (Fowler and Thomashow, 2002; Chinnusamy et al., 2003).

The concentrations of CRLK1$_a$ (**Figure 2B**), MPK6P (**Figure 2C**), ICE1P (**Figure 2D**), MPK4P (**Figure 2F**), HOS1n (**Figure 2G**), *CBF3* mRNA (**Figure 2H**), and *ZAT12* mRNA (**Figure 2I**) all increase after the cold stress, following the transient increase in Ca^{2+} (**Figure 2A**). Because the Ca^{2+} influx is transient, we see that the response of the network is also transient and the variables—including Ca^{2+}—return to the steady state that prevailed prior to the cold stress. What is striking is the difference in time scales: while the Ca^{2+} signal rises in a few seconds and is over in a few minutes, the increase in *CBF3* mRNA peaks a few hours after the temperature drop and remains elevated for some 10 h. This represents a most significant result of our computational model, which compares well with the experimental observations, as indicated in **Figure 2**.

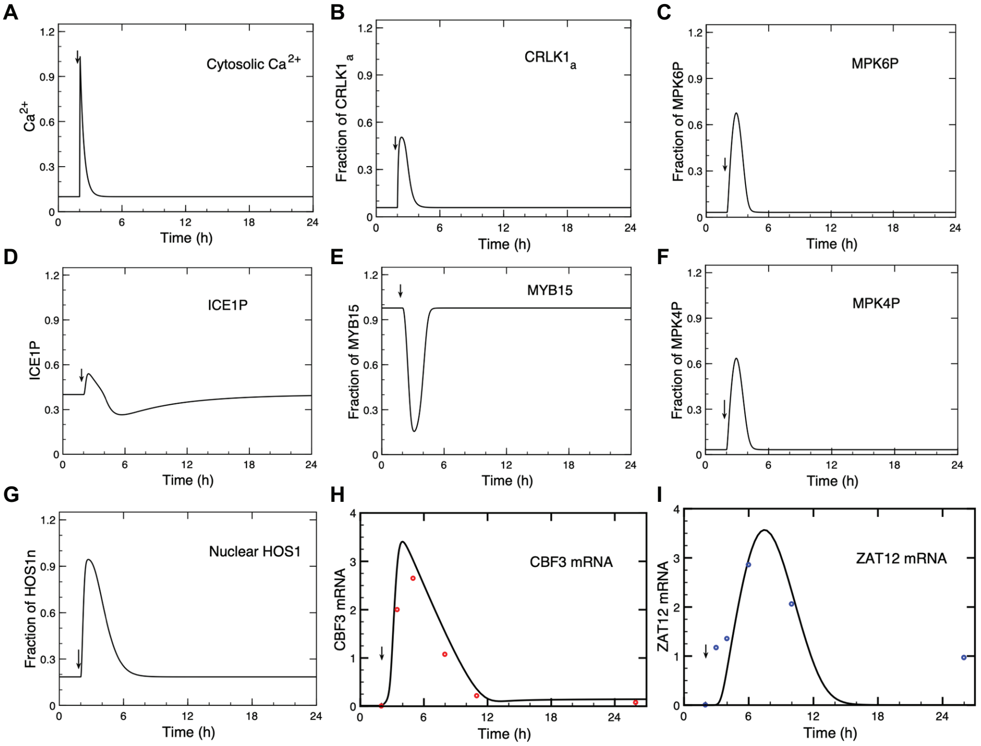


FIGURE 2 | Dynamics of the cold response pathway after a cold stress. **(A)** The cold stress produces a transient increase in Ca^{2+} at time = 2 h. This brief Ca^{2+} pulse induces a transient change in the successive steps of the cascade: CRLK1a **(B)**, MPK6P **(C)**, ICE1P **(D)**, MYB15 **(E)**, MPK4P **(F)**, HOS1n **(G)**, CBF3 mRNA **(H)**, and ZAT12 mRNA **(I)**. **(H,I)** Comparison of theoretical predictions (solid lines) for the time course of CBF3 mRNA and ZAT12 mRNA with experimental data for CBF3 (red dots) and ZAT12 (blue dots) redrawn, respectively, from **Figure 1** in Zarka et al. (2003) and from **Figure 7A** in Fowler and Thomashow (2002), by means of the software ImageJ for Image Processing and Analysis (https://imagej.nih.gov/ij/). Time is increased by 2 h when redrawing the experimental points from the original publications, because in the simulations the cold stress is given at time = 2 h (to determine the steady-state levels before the cold signal is applied) instead of 0 h in the experiments. The results are obtained by numerical integration of kinetic equations (4)–(17), by means of the program XPP (http://www.math.pitt.edu/~bard/xpp/xpp.html), for a Ca^{2+} pulse generated according to eqs. (1a–c), while the associated changes in the fraction of activated calmodulin (CaM) are determined according to eq. (3), as described in Sections 1 and 2 in **Supplementary Material**. Parameters are defined in **Supplementary Table 2** where their numerical values are listed.

Long after the concentration of Ca^{2+} has decreased, the concentrations of the other state variables of the model also decrease but this occurs on a time scale that becomes longer and longer as the effect of the brief transient signal of Ca^{2+} propagates through the successive reaction steps of the network.

As seen in **Figure 2E**, in contrast to the other variables plotted in the figure, the concentration of MYB15 decreases in response to the cold signal, before returning to its steady state level as Ca^{2+} decreases. When the amount of MYB15 reaches its trough, the inhibition of CBF3 decreases. As the amount of ICE1P concomitantly peaks (**Figure 2D**), the level of *CBF3* mRNA rises. Both the decrease in the unphosphorylated form of MYB15 and the increase in the phosphorylated form of ICE1 are brought about by their phosphorylation by MPK6P, which follows the activation of this MAP kinase by the rise in Ca^{2+} (**Figure 2A**).

To better grasp how the dynamics of the cold response pathway unfolds in the course of time, we plot in **Figure 3** on an enlarged time scale the evolution of Ca^{2+}, $CRLK1_a$, ICE1P, HOS1n, MPK6P, MPK4P, MYB15, CBF3 mRNA, and ZAT12 mRNA during the first 2 h (A) and 4 h (B), after the cold stress. The Ca^{2+} pulse, triggered by the shift to cold temperatures, starts at t = 2 h. Here, plotting the different curves on the same graph allows us to clarify the sequence of activation steps that leads to *CBF3* expression in the pathway. The state variables start from their steady state levels reached before the cold stress and, after the cold signal is given, increase or decrease in response to the Ca^{2+} pulse. This pulse is followed, successively, by the peaks in $CRLK1_a$, ICE1P, HOS1n, MPK6P, MPK4P, then by the trough in MYB15, and the peaks in *CBF3* mRNA and *ZAT12* mRNA. Some of the variables, such as Ca^{2+}, $CRLK1_a$, MYB15, ICE1P, MPK6P, and MPK4P, return more rapidly to their steady state than other variables which take more time to do so (see **Figure 3B**). The return to steady state of *CBF3* mRNA, followed by *ZAT12* mRNA (induced by CBF3) takes more time, because the degradation of *CBF3* mRNA proceeds at a relatively slower rate at cold temperatures.

The predicted evolution of ICE1P appears to be biphasic. First ICE1P rises and peaks in advance of its upstream variables MPK6P and MPK4P, and later decreases below steady state, due to the rise in HOS1n which promotes the degradation of ICE1P. After MPK6P, MPK4P and HOS1n reach their steady state, ICE1P begins to rise again after t = 6 h (see **Figure 2D**), because of the input due to constant synthesis of the ICE1 protein at a rate v_{s1} in eq. (10), and its conversion, activated by MPK6P, into the phosphorylated form ICE1P. This slow rise continues until ICE1P reaches its steady state. We included a term for the synthesis of ICE1 to avoid the depletion of the protein and to ensure the existence of a steady state, because we take into account the degradation of ICE1P, which is enhanced by HOS1n.

The results of **Figures 2** and **3** illustrate how the model schematized in **Figure 1** provides us with a detailed, integrated view of the chain of molecular events that leads from the cold-induced Ca^{2+} pulse to the synthesis of *CBF3* mRNA.

Dynamics of the Model Under Single Cold Shock: Response to a Train of Ca^{2+} Oscillations

In guard cells, a cold stress elicits a series of small, repetitive [Ca^{2+}] transients, with an amplitude of about 125 nM and a period of about 150 s, associated with stomatal closure (Allen et al., 2000, 2001). The cold stress regulates the flow of Ca^{2+} into the cytosol, which primes the apoplast for oscillatory cycles of Ca^{2+} release. To model these oscillations so as to determine their effect on the dynamics of the cold response pathway, we assume in the model that the influx of Ca^{2+} from the apoplast is proportional to parameter β (red curve in **Figures 4A,B**) which measures the magnitude of the stimulus in the two-variable model for signal-induced Ca^{2+} oscillations (Goldbeter et al., 1990; Dupont et al., 1991).

In order to take into account the transient nature of the changes in Ca^{2+}, we consider an instantaneous increase in β followed by an exponential decrease, which results in a transient

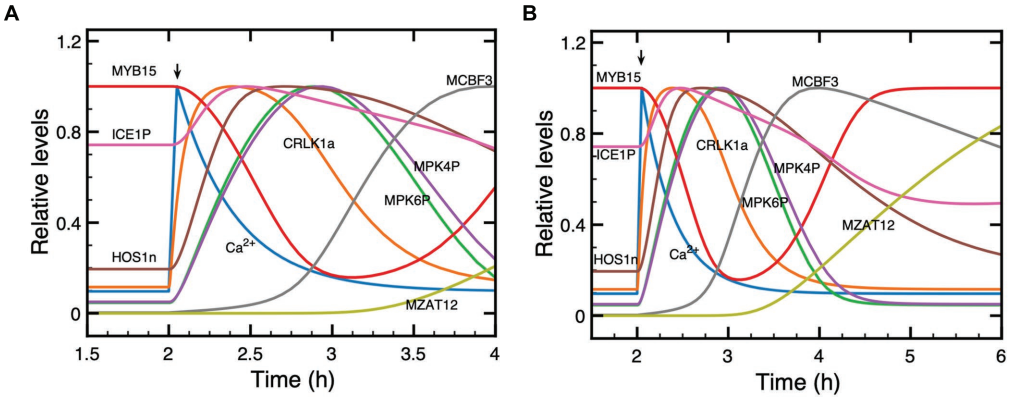


FIGURE 3 | Dynamics of the cold response pathway during the first 2 h **(A)** or 4 h **(B)** after a cold stress. Shown is the time evolution of Ca^{2+}, $CRLK1_a$, ICE1P, HOS1n, MPK6P, MPK4P, MYB15, CBF3 mRNA, and ZAT12 mRNA. The cold signal elicits at t = 2 h a transient Ca^{2+} pulse (blue curve). Before the pulse, all state variables are at their stable steady state. To distinguish more clearly the relative positions of the different curves, each plotted state variable was normalized through division by the maximum it reaches in the course of time, since absolute levels would make it difficult to follow the order of peaks and troughs. Parameter values are the same as in **Figure 2** where the time evolution is shown separately for each variable on a longer time scale.

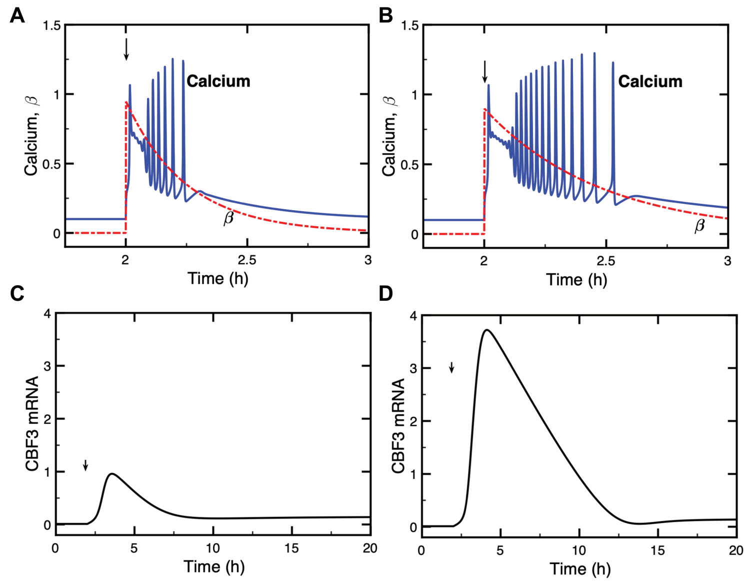


FIGURE 4 | Dynamics of the model for the cold response pathway when the cold stress elicits a train of high-frequency Ca^{2+} oscillations. Panels **(A,B)**: the single cold shock elicits a train of Ca^{2+} pulses, which contains a few **(A)** or a larger number **(B)** of Ca^{2+} spikes. Panels **(C,D)** show the evolution of *CBF3* mRNA corresponding to panels **(A,B)**, respectively. After an initial increase at time t_p = 2 h, parameter β decays exponentially according to the equation $\beta = \beta_f \exp\left[-a\left(t - t_p\right)\right]$ with β_f = 95% and a = 4 in **(A)** and β_f = 90% and a = 2.1 in **(B)**. The curves for *CBF3* mRNA are obtained by numerical integration of the kinetic equations (4)–(17) (see **Supplementary Material**, Section 2) by means of XPP (http://www.math.pitt.edu/~bard/xpp/xpp.html), while the oscillatory Ca^{2+} dynamics and the associated changes in the fraction of activated *CaM* are determined as described in **Supplementary Material** Section 1.2, according to eqs. (2) and (3), with parameter values from **Figure 2** in Dupont et al. 1991 multiplied by 25 to obtain the appropriate time scale.

train of Ca^{2+} oscillations (blue curves in **Figures 4A,B**). The expression of *CBF3* mRNA in response to this train of transient spikes is determined by numerical simulations, which demonstrate the effect of a train of Ca^{2+} oscillations (**Figures 4C,D**). The simulations of the model indicate that the longer the duration of Ca^{2+} oscillations, the higher the magnitude of *CBF3* mRNA accumulation in response to cold stress. The duration of the train of Ca^{2+} spikes depends on the rate of decrease in β, after the initial increase in this parameter brought about by the decrease in temperature.

What distinguishes the dynamics of the cold response to a single Ca^{2+} pulse or to a train of Ca^{2+} oscillations? To address this question, we compare the two situations in **Supplementary Figure S1**. The longer a train of oscillations, the larger the number of Ca^{2+} peaks and, therefore, the greater the magnitude of the Ca^{2+} signal acting on the cold response pathway. In panels A and C of **Supplementary Figure S1**, we numerically computed the area under the curve (AUC) corresponding to the oscillations in Ca^{2+} shown in **Figures 4A,B**; the corresponding *CBF3* mRNA peaks elicited by these oscillations are shown in **Figures 4C,D**, and redrawn in panels B and D in **Supplementary Figure S1**. Then, we generated single pulses of Ca^{2+} with comparable AUCs, as shown in panels E and G in **Supplementary Figure S1**. The peaks in *CBF3* mRNA produced in response to these pulses closely match those produced in response to Ca^{2+} oscillations with similar AUCs (compare in **Supplementary Figure S1** panels F and H with panels B and D, respectively). It appears, therefore, that what governs the CBF3 response is the amount of Ca^{2+} liberated by the cold stress in the form of a single Ca^{2+} pulse or a train of Ca^{2+} oscillations.

Effects of Mutations in the Cold Stress Pathway

The model of the cold response pathway allows us to predict the effect of mutations in the various elements of the network and to compare the results with available experimental data. We determine the expression of *CBF3* as the output when the levels of ICE1, HOS1, MYB15, and ZAT12 are altered as a result of mutations or overexpression. For the proteins for which the total amounts are assumed to be constant, to represent such mutations, we reduce the total amounts of MYB15 to 10% of the total amounts in wild type (WT), while we increase 5-fold the total WT concentration for HOS1 overexpression (OX) mutants. Moreover, the rate of synthesis of ICE1 is reduced to 20% to represent the effect of mutations, while the rate of synthesis of ZAT12 is increased 5-fold to represent the effect of overexpression. We compare the predictions of the model (solid lines, in black for the WT and in blue for the mutants) with experimental data (red dots) in **Figure 5**.

In **Figure 5A**, the *ice1* mutation blocks the CBF3 transcripts, in agreement with experimental observations (Chinnusamy et al., 2003), while overexpression of ICE1 (ICE1-OX) significantly enhances the expression of *CBF3* (**Supplementary Figure S2**) in agreement with the observed enhancement in cold tolerance (Chinnusamy et al., 2003; Lee et al., 2005). In contrast to

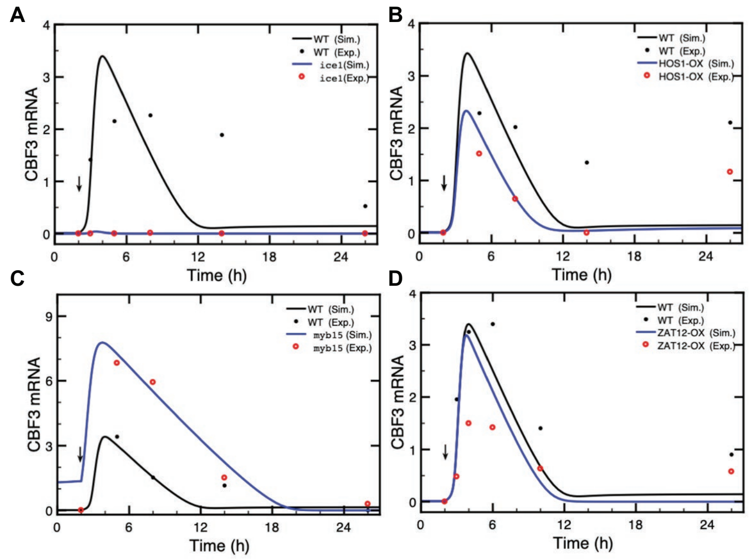


FIGURE 5 | Effect of loss-of-function mutations or overexpression (OX) of genes in the cold response pathway. The expression of CBF3 is determined in the model after a cold shock given at time = 2 h (vertical arrow) in the form of a Ca^{2+} pulse. In each panel, we compare the simulated CBF3 mRNA levels in wild type (WT, black curves), and mutants (blue curves) with the CBF3 mRNA level determined experimentally in WT (black dots) and in mutants (red dots) in **(A)** the *ice1* mutant, **(B)** the overexpression of the HOS1 gene (HOS1-OX), **(C)** the *myb15* mutant, and **(D)** the overexpression of ZAT12 (ZAT12-OX). In each panel, the time at which experimental points were taken is increased by 2 h because in the simulations the cold stress is given at time = 2 h instead of 0 h in the experiments (thus a point measured at 6 h is plotted at 8 h in **Figure 4B**, to allow the comparison between the simulations and the experimental results). The theoretical curve (black line) obtained by simulations for the WT is the same in all panels—the changed scale in panel **(C)** is due to higher levels of *CBF3* mRNA in the *myb15* mutant—while the experimental points for the WT (black dots) are taken from different publications, as mentioned hereafter. In **(A)**, the value of ICE1 synthesis rate in the *ice1* mutant is 0.036, compared to 0.18 in WT (see **Supplementary Table 2** for units of the parameters). Experimental data are redrawn, by means of the software ImageJ for Image Processing and Analysis (https://imagej.nih.gov/ij/), from **Figure 1C** in Chinnusamy et al. (2003). In **(B)**, the total amount of HOS1 protein in HOS1-OX is 10, vs. 2 in WT. Experimental data for CBF3 expression in HOS1-OX are redrawn from **Figure 7B** in Dong et al. (2006). The peak in CBF3 expression at 24 h (corresponding to time = 26 h in **Figure 4B**) is also observed in the wild type; see **Figure 7B** in Dong et al. (2006). This increase is not accounted for by the model (see "Discussion" section). In **(C)**, the total amount of MYB15 protein in the *myb15* mutant is 0.2, vs. 2 in WT. Experimental data are redrawn from **Figure 7C** in Agarwal et al. (2006). In **(D)**, the synthesis rate is 11 in ZAT12-OX, vs. 2.2 in WT. Experimental data are redrawn from Figure 10A in Vogel et al. (2005). Parameter values are listed in **Supplementary Table 2** except for $vs1$ = 0.036 nM/h, $vs3$ = 11 nM/h (overexpression of ZAT12), $HOS1_t$ = 10 nM, and $MYB15_t$ = 0.2 nM. The results are obtained by numerical integration of kinetic equations (1) and (4)–(17) from **Supplementary Material** Sections 1 and 2, using the XPP program (http://www.math.pitt.edu/~bard/xpp/xpp.html).

ICE1, HOS1 is a negative regulator of low temperature-responsive gene transcription. HOS1 accumulates in the nucleus in response to low temperature treatments. The overexpression (HOS1-OX) leads to the degradation of ICE1P and represses the expression of *CBF3* mRNA (Lee et al., 2001; Dong et al., 2006). **Figure 5B** indicates that in the HOS1-OX mutant, the synthesis of *CBF3* mRNA is accordingly attenuated.

In contrast to ICE1, MYB15 is a repressor of *CBF3*. Following a cold stress, MYB15 is inactivated through phosphorylation, which leads to a rise in *CBF3* transcription. Overexpression of MYB15 results in the reduced expression of *CBF* genes whereas its loss-of-function leads to increased expression of *CBF* genes in the cold (Agarwal et al., 2006). **Figure 5C** and **Supplementary Figure S3** illustrate the enhanced *CBF3* expression in the *myb15* mutant.

Finally, we turn to ZAT12, which dampens the expression of the CBF cold response pathway (Vogel et al., 2005). The overexpression of ZAT12 represses the level of *CBF3* mRNA when the level of ZAT12 protein increases above a threshold value. The expression of *ZAT12* is itself induced by CBF3. **Figure 5D** shows that during the first 2 h after the cold stress, the level of *CBF3* mRNA is nearly the same as in the WT because the ZAT12 protein has not yet accumulated, through induction by CBF3. Later, the higher level of the ZAT12 protein in ZAT12-OX accelerates the decrease in CBF3 because of the negative feedback exerted by ZAT12 on CBF3. As to the 14-3-3 protein, which phosphorylates CBF3, mutations in 14-3-3 reduce this phosphorylation and, hence, the expression of *ZAT12* mRNA (data not shown). The situation in regard to *CBF3* mRNA is, therefore, similar to that in *ZAT12* mutant.

Dynamics of the Model Under Successive Cold Stresses

We focused in the previous sections on the situation where the temperature drops rapidly and produces a single transient increase in Ca^{2+} or a train of Ca^{2+} spikes. However, the temperature fluctuates seasonally, daily, as well as on a short time scale due to stochastic fluctuations. The time scale over which Ca^{2+} changes occur varies greatly, from rapid increases within seconds (Knight et al., 1996) to slower, 24 h daily rhythms (Dodd et al., 2006; Ruiz et al., 2018, 2020). Moreover, the CBF3 transcripts, in response to an instantaneous cold stress, vary during the day with a peak at about 8 h after dawn (ZT8) and with a trough at about ZT20. Cold-induced changes in *CBF3* transcription are themselves gated by the circadian clock (Fowler et al., 2005; Bieniawska et al., 2008; Nakamichi et al., 2009; Dong et al., 2011; Jiang et al., 2020), an effect to which we will return in "Circadian Gating" section below.

The dynamics of the cold response pathway under successive cold stresses has been studied experimentally by Zarka et al. (2003). Their observations allow us to compare the predictions of the model with the results they obtained in such conditions. We represent the warm-cold cycle by a series of identical Ca^{2+} pulses and take into account the increased rate of *CBF3* mRNA degradation at higher temperatures. The time evolution of *CBF3* mRNA under different warm-cold cycles is presented in **Figure 6**. We tested 3 cycles characterized by a period of 3 h (90 min cold followed by 90 min warm, as in the experiments of Zarka et al., 2003; **Figure 6A**), 6 h (**Figure 6B**), and 24 h (**Figure 6C**), respectively. In each warm-cold cycle, we assume that Ca^{2+} increases transiently at the beginning of the cold interval and that the degradation rate of *CBF3* mRNA rises 10-fold in the warm phase.

Numerical simulations indicate that the *CBF3* mRNA level, which increases in response to the cold stress and diminishes due to both the end of the Ca^{2+} pulse and the enhanced degradation at higher temperatures, undergoes periodic variations, in phase with the warm-cold cycles. The level of *CBF3* mRNA under 3 h-warm-cold cycles displays desensitization of the cold response (Zarka et al., 2003), since the amplitudes of *CBF3* mRNA in each cycle gradually declines (**Figure 6A**). Under natural circumstances, the temperature fluctuates daily with, typically 12 h warm and 12 h cold. **Figure 6C** shows that the amplitude of *CBF3* mRNA then remains the same in each cycle. This is due to the decay of downstream inhibitors, such as ZAT12, in the course of time, which allows CBF3 to return to high levels. In **Figure 6B**, the amplitudes of *CBF3* mRNA in 6 h cycles drops but significantly less compared to the 3 h cycles. This suggests that the time to restore the transcription rate at maximum level is about 12 h. The corresponding value shown by the data of **Figure 2B** in Zarka et al. (2003) is between 8 and 24 h.

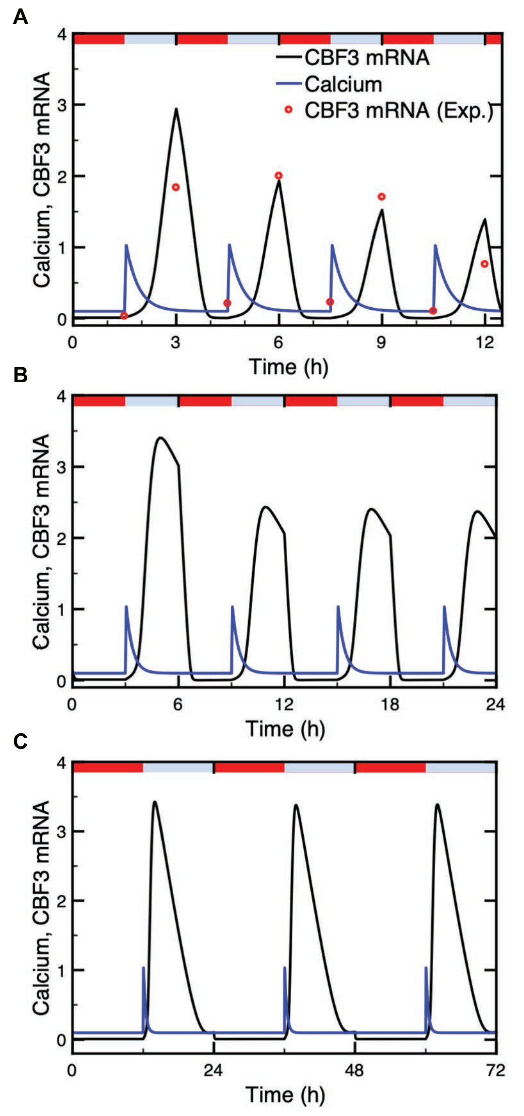


FIGURE 6 | Dynamics of the cold response pathway subjected to repetitive cold stresses. Shown is the CBF3 mRNA (black curve) under identical Ca^{2+} pulses (blue curve) administered at intervals of 3 h **(A)**, 6 h **(B)**, and 24 h **(C)**. The Ca^{2+} pulses in the three panels have the same duration. Identical Ca^{2+} pulses are assumed to be elicited at the beginning of each cold phase, which is followed by a warm phase of similar duration. The duration of these phases is equal to 90 min **(A)**, 3 h **(B)**, or 12 h **(C)**. Repetitive cold stimuli are assumed to produce identical Ca^{2+} spikes. The experimental data (red points) are redrawn from **Figure 2A** in Zarka et al. (2003). The red bars and blue bars on top of each panel represent warm (20°C) and cold (4°C) phases, respectively. Parameter values are listed in **Supplementary Table 2**. The results are obtained by numerical integration of eqs. (1) and (4)–(17), and by means of eq. (3) for the determination of *CaM* (see **Supplementary Material** Sections 1 and 2), using the XPP program. The degradation rate of CBF3 mRNA is 0.55 and 5.5 in cold and warm phases, respectively.

Putative CBF3 Oscillations Resulting From Negative Feedback Exerted by ZAT12

The investigation of the dynamics of the model for the plant response to cold stress has revealed, unexpectedly, the possibility of sustained oscillations in the expression of *CBF3*. This phenomenon, which occurs spontaneously for a restricted set of parameter values, does not need to be triggered by the Ca^{2+} influx that occurs upon decreasing temperature to a low level. Such oscillations are shown in **Figures 7A,B** for the mRNA and protein levels, respectively, both for CBF3 and ZAT12. Projecting the trajectory onto a phase plane formed by *CBF3* mRNA and *ZAT12* mRNA indicates that this periodic behavior corresponds to the evolution toward a closed curve known as limit cycle (**Figure 7C**). The conspicuous property of the limit cycle is that it can be reached regardless of initial conditions and, therefore, represents a very robust mode of oscillations. The other state variables upstream in the plant cold response pathway (see **Figure 1**) evolve to a stable steady state after the cold stress. Thus, the oscillations in CBF3 and

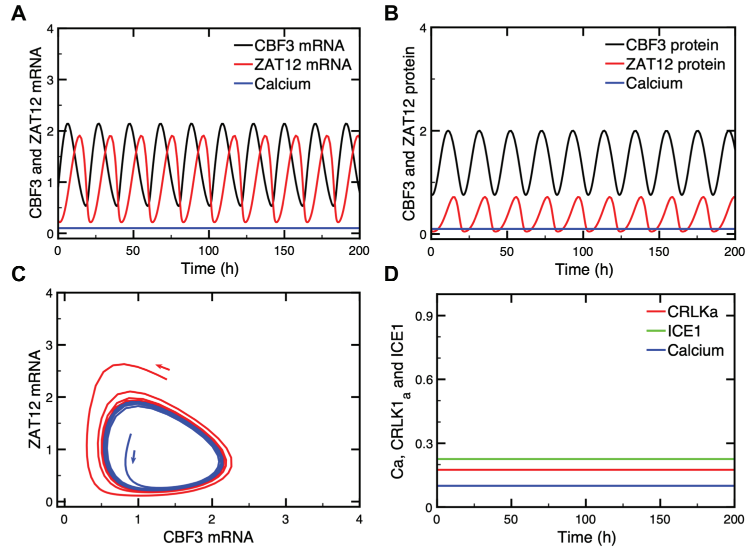


FIGURE 7 | Oscillations in CBF3 expression in the absence of a cold shock. In some conditions, oscillations can occur spontaneously in some state variables of the cold response pathway, such as CBF3 and ZAT12, both in their mRNAs **(A)** and protein levels **(B)**. When the levels of CBF3 and ZAT12 mRNAs are plotted against each other in the course of time, the oscillations correspond to the evolution toward a closed curve known as a limit cycle **(C)**. This limit cycle can be reached starting from different initial conditions, as shown by the trajectories in blue and red in **(C)**. By contrast, other state variables located upstream of CBF3 in the cold response pathway do not oscillate, as shown in **(D)** for CRLK1a (red curve) and ICE1 (green curve) which remain at a stable steady state, in the absence of cold stress. The Ca^{2+} level (blue curve) accordingly remains also at a stable steady state level. The curves are obtained by numerical integration of eqs. (1), (3), and (4)–(17) from **Supplementary Material** Sections 1 and 2. Parameter values are listed in **Supplementary Table 2** except for the following parameters: $K_{a2} = 0.55$, $K_{I2} = 0.8$, $K_{I3} = 0.2$, $K_{d3} = 1.8$, $k_{s1} = 0.25$, $v_{d5} = 3.8$, $v_{s3} = 3.5$, $K_{a3} = 0.8$, $v_{d6} = 3.2$, $k_{s2} = 0.55$, $v_{d7} = 3.65$, and $k_{d7} = 0.1$ (see **Supplementary Table 2** for the definitions and units of these parameters).

ZAT12 are not observed in other proteins involved in the cold response pathway, as shown in **Figure 7D**.

What is the origin of such CBF3 oscillations? Negative feedback represents a well-known source of oscillatory behavior in biochemical systems at the cellular level, as exemplified by the molecular mechanism of the circadian clock, which involves auto-regulatory negative feedback loops on gene expression in a variety of organisms, including *Drosophila*, mammals, and plants (Goldbeter, 1995; De Caluwé et al., 2016). Many other examples of oscillations originate from regulatory feedback processes at the cellular level (see Goldbeter, 1996; Goldbeter et al., 2012, for reviews). Here, a negative feedback is exerted by ZAT12 on the expression of *CBF3*. Indeed, the expression of *ZAT12* is induced by CBF3 while ZAT12 represses the expression of *CBF3*. This negative feedback regulation is responsible for the oscillatory expression of CBF3, which is characterized by a period of about 32 h for the parameter values considered in **Figure 7**. The characteristics of these oscillations are further explored in **Supplementary Material**, Section 4 and in **Supplementary Figure S4**.

That the mechanism of the oscillations relies on the negative feedback exerted by ZAT12 on *CBF3* expression is shown by the observation that while CBF3 and ZAT12 oscillate, the other state variables located upstream in the cold response pathway (see **Figure 1**) do not display oscillations and remain at a stable steady state, as illustrated in **Figure 7D**. This difference is due to the fact that the regulation exerted by ZAT12 at the last step of the pathway schematized in **Figure 1** does not affect the synthesis or degradation of the state variables involved in the preceding steps of the network.

What are the conditions that could favor the onset of oscillations in CBF3 expression? The stability diagram in **Supplementary Figure S4D** suggests that oscillations might be induced by decreasing the rate of degradation of *CBF3* mRNA, v_{d3}, when the system moves first from the stable steady state P_2 to the stable steady state P_3 (in which the level of *CBF3* mRNA is higher, as can be seen in **Supplementary Figures S4E,F**, because of the reduced rate of degradation) and then to the unstable steady state P_3 around which sustained oscillations occur. The same diagram in **Supplementary Figure S4D** shows that oscillations could also result from increasing the maximum rate of *CBF3* expression v_{s2} at a given value of v_{d3}. The model predicts, however, that other parameter changes may also induce oscillations, for example—as suggested by the diagram in **Supplementary Figure S4B**—by increasing (on the left of this diagram) or decreasing (on the right of this diagram) the maximum rate of ZAT12 mRNA degradation, v_{d6}, or by increasing the ratio k_9/v_5 which measures the ratio of phosphorylation and dephosphorylation rates of ICE1. Increasing this ratio favors the rise in ICE1P, which enhances the expression of *CBF3* (see scheme of the model in **Figure 1**). In conclusion, all parameters which directly or indirectly enhance the formation of CBF3, by increasing its synthesis or decreasing its degradation, favor the onset of oscillations, provided that the level of ZAT12 is in an appropriate range, as shown in **Supplementary Figure S4B**.

Circadian Gating

Many aspects of plant physiology and metabolism are controlled by the circadian clock. Unsurprisingly, the cold response was also shown to be under the circadian control (Fowler et al., 2005; Dong et al., 2011). Circadian rhythms are generated at the cellular level by a dozen of clock genes, which form a network of interlocked regulatory feedback loops (Nohales and Kay, 2016; McClung, 2019). This genetic clock is entrained by the external light-dark cycle and controls the expression of many genes, including *CBF* genes (Fowler et al., 2005; Dong et al., 2011). Moreover, the circadian phase at which plants were transferred to low temperature affects the level at which *CBF* transcripts accumulate upon cold stress (Fowler et al., 2005). That the induction of *CBF* genes is gated by the circadian clock was corroborated by subsequent experiments (Dong et al., 2011; Keily et al., 2013).

To account for this circadian gating in our model, we simulated the circadian clock signal by a sine function (see **Supplementary Material**, Section 3). This circadian variable, which could represent the level a circadian clock protein (e.g., CCA1/LHY), acts at the level of *CBF3* expression, as expressed by eqs. (18)–(19); (see **Supplementary Material**, Section 3). In **Figure 8A**, we show the circadian variation in *CBF3* expression controlled by the circadian clock in the absence of cold stress. Although the LD cycle is not modeled explicitly in this work, we add for clarity the 12 h:12 h LD cycle on top of the panel, with a phase set so that the peak in *CBF3* mRNA occurs at ZT 8, as observed experimentally (Dong et al., 2011).

Our simulations show that, depending on the time at which the Ca^{2+} pulse triggered by the cold stress is applied, *CBF3* expression can be strongly induced (**Figure 8B**) or not (**Figure 8C**). The responsiveness to cold stress is shown in **Figure 8D** as a function of the timing of the Ca^{2+} pulse initiated by the cold stress. These results support the view that the circadian clock limits the cold response to a certain window of time.

DISCUSSION

The aim of this study was to build a detailed computational model of the pathway responsible for the plant response to cold stress, based on available experimental observations. Our work provides a theoretical framework that we used to address the dynamics of this important regulatory network in a variety of conditions. The model takes into account the results from numerous biochemical and genetic studies which showed how a cold stress activates a signaling cascade leading to the expression of *COR* genes, which enable plants to tolerate freezing temperatures (Thomashow, 1999; Gilmour et al., 2004; Hannah et al., 2006; Lissarre et al., 2010). The cold stress triggers a

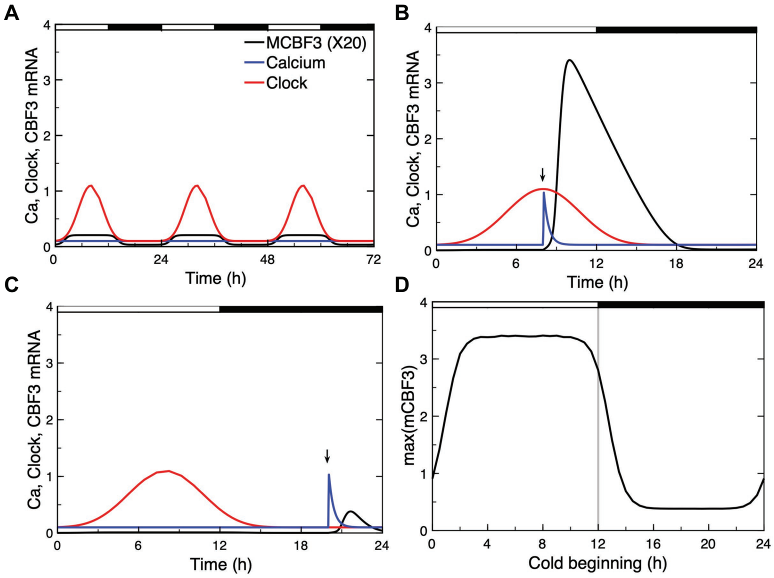


FIGURE 8 | Circadian gating of the cold response. Depending on the time of the day at which the cold stress is applied, CBF3 expression can be strongly induced or not. **(A)** In the absence of cold stress, the Ca^{2+} level (blue curve) remains at a stable steady state, while CBF3 mRNA undergoes small amplitude oscillations (black curve) driven by the circadian clock (red curve). The 12 h light (L)-12 h dark (D) LD cycle is represented on top of the figure in such a way that the maximum in CBF3 mRNA occurs at ZT8, i.e., 4 h before the L to D transition, as observed experimentally (Dong et al., 2011). For the sake of clarity, the level of CBF3 mRNA in **(A)** is multiplied by 20. **(B)** When the cold stress, resulting in a brief Ca^{2+} pulse, is applied at a phase when CBF3 driven by the circadian clock is near its maximum, the cold stress results in a large-amplitude peak in CBF3 produced by the plant cold response pathway. **(C)** By contrast, when the Ca^{2+} pulse is applied at a phase when CBF3 is away from its maximum, the cold stress fails to elicit the activation of the plant cold response pathway. **(D)** Plotting the maximum of the CBF3 response as a function of the time at which the same Ca^{2+} pulse is applied shows that the circadian clock defines a window of time during which a significant cold response can be achieved. The curves are obtained as described for **Figure 2**, for the parameter values listed in **Supplementary Table 2**. The circadian oscillation that controls CBF3 expression is generated according to eqs. (18) and (19) in **Supplementary Material** Section 3, with $K_c = 0.15$ and $C_0 = 0.01$.

signal in the form of a brief Ca^{2+} spike, initiating a cascade of biochemical reactions that eventually lead to a rise in the factor CBF3, which induces the expression of *COR* genes after several hours. As shown in **Figures 2** and **3**, the model allows us to follow how the initial signal is transduced through successive steps which eventually produce the rise in CBF3. Experimentally, the plant response to cold stress is measured by the level of *CBF3* mRNA transcripts which, in the wild type, peaks between 1 and 6 h after cold stress, according to different experiments (Fowler and Thomashow, 2002; Chinnusamy et al., 2003). In the model, we consider that the brief Ca^{2+} pulse represents the initial response to the cooling stimulus. Remarkably, this initial Ca^{2+} pulse, which rises in seconds and is over in minutes, results in the sequential activation of various protein kinases, and finally in the transient expression of genes over a much longer time scale, of the order of hours.

The model allows us to follow *in silico* the detailed chain of biochemical events that leads from the cold-induced Ca^{2+} signal to *CBF3* expression. As done in **Figures 2, 3**, we can use the computational model as a magnifier to enlarge the details of the time evolution of the different variables in the model so as to disentangle their time course in the minutes and hours that follow the cold stress. Obtaining experimentally such detailed comparative information would be extremely valuable. The predictions of the computational model provide us with a framework that may prove useful to this end.

Because of the many regulatory interactions that link the biochemical variables in the cold response pathway schematized in **Figure 1**, the detailed sequence of events leading to *CBF3* expression is sometimes difficult to predict in the absence of a computational model. Thus, in the chain of sequential activations, as shown in **Figure 3**, once activated by CaM CRLK1a activates the kinases MPK6 and MPK4 which exert antagonistic effects on ICE1P. Moreover, CRLK1 also activates HOS1 which controls the degradation of ICE1P. The time course of ICE1P is, therefore, controlled by the combined effects of MPK6, MPK4, and HOS1. Similarly, the time evolution of *CBF3* mRNA is activated by ICE1P and inhibited by MYB15 and ZAT12, the latter being induced by CBF3. The model allows us to predict the time course of all these variables as it takes into account and integrates simultaneously all regulatory interactions in the pathway. It might be difficult to reach such a comprehensive, global view of the dynamics of the cold response pathway without a computational approach.

Rather than eliciting a single Ca^{2+} spike, the cold stress sometimes triggers a train of high-frequency Ca^{2+} oscillations (Allen et al., 2001). We determined the effect of such an oscillatory signal on the response of the model for the cold response pathway using a generic model for Ca^{2+} oscillations based on the Ca^{2+}-induced Ca^{2+} release mechanism (Goldbeter et al., 1990; Dupont et al., 1991). As shown in **Figure 4**, because it occurs on a much longer time scale the expression of *CBF3* integrates the high-frequency Ca^{2+} spikes without reflecting each of them. The amplitude of the response thus increases with the number of spikes and, hence, with the duration of the Ca^{2+} oscillatory signal.

We used the model to predict the effect of mutations or overexpression of various genes in the cold response pathway. Every step in the cascade leading from the initial Ca^{2+} pulse to *CBF3* expression is capable of affecting the response to cold stress. As shown in **Figure 5** and in **Supplementary Figures S2**, **S3**, the predictions of the model agree with experimental observations as to how the expression of *CBF3* is affected by mutations in or overexpression of various genes such as those coding for ICE1, HOS1, MYB15, or ZAT12. In **Figure 4B**, established for HOS1-OX, the experimental data for *CBF3* expression (red dots) display a peak at 24 h (corresponding to time = 26 h in **Figure 4B**, as explained in the legend to **Figure 4**). This peak is also observed in the wild type; see **Figure 7B** in Dong et al. (2006). Such increase, which is not accounted for by the simulations of the model is, however, not always observed in the experiments; see, for example, **Figure 2H** in Zarka et al. (2003). It might perhaps be due to another mechanism, such as circadian gating, considered in “Circadian Gating” section.

Experimental observations indicate that *CBF3* transcription displays desensitization and resensitization in response to successive cold stresses (Plieth et al., 1999; Zarka et al., 2003). Both for Ca^{2+} and *CBF3* transcripts, when the decreasing temperature stimuli are repeated, the amplitudes of the $[Ca^{2+}]$ spikes and *CBF3* expression levels decrease (Plieth et al., 1999). The dependence on past low temperature episodes that the plant previously encountered indicates that the $[Ca^{2+}]$ spikes and *CBF3* transcript response to cooling or in response to a stepwise decrease of the temperature can be attenuated (Knight et al., 1996; Zarka et al., 2003). The cold-sensing mechanism resensitizes after 8–24 h at warm temperatures (Zarka et al., 2003). The results of numerical simulations of the model indicate that even in the case where the system is subjected to a series of identical Ca^{2+} pulses corresponding to repetitive alternations of cold and warm phases, the cold response pathway is capable of displaying by itself desensitization (**Figure 6A**). The phenomenon attenuates (**Figure 6B**) and finally disappears (**Figure 6C**) when the interval between successive cold phases increases sufficiently.

As with all models involving a large number of state variables and parameters, a major question arises in regard to parameterization, given that quantitative information on many parameters in such models is often scarce, if not unavailable. As exemplified by more complex models involving even greater numbers of state variables and parameters, such as those proposed for the circadian clock (Leloup and Goldbeter, 2003) or the mammalian cell cycle (Gérard and Goldbeter, 2009), the computational models for cellular regulatory pathways remain useful for studying the mechanism and dynamics of these processes at the molecular and systems levels, and sometimes to predict new, counter-intuitive modes of qualitative behavior, even though precise parameter values are lacking in these models. One way of dealing with the issue of parameterization is to study numerically the dynamic behavior of the model over large ranges of parameter values. In addition, a sensitivity analysis may be performed to investigate the range of parameter values in which certain experimental observations can be reproduced (see, for example, Leloup and Goldbeter, 2004; De Caluwé et al., 2016).

As explained in detail in Section 4 of **Supplementary Material**, both approaches were followed in this study of the model for the cold response pathway, which contains 14 state variables and 77 parameters, not counting those related to Ca^{2+} and CaM. Most of these parameter values were not (yet) characterized experimentally. For example, kinetic studies have yet to be performed to yield values for the various rates of phosphorylation or dephosphorylation, for the Michaelis constants that characterize these enzyme reactions, for the activation or inhibition constants that measure the regulatory interactions, and for the total concentrations of the different molecular species involved in the pathway. For all these parameters, we made a semi-arbitrary choice, in a physiologically reasonable range, as explained in Section 4.1 of **Supplementary Material**.

The choice of parameter values was guided by the constraint of matching the observed characteristics of the *CBF3* expression, which is the major response measured in the experiments, in regard to its amplitude, timing, and half-width. This was achieved by trial and error, i.e., by performing simulations for a large set of parameter values and selecting a set that produced (1) a *CBF3* mRNA peak in a range extending from 1 to 6 h after the cold stress, as observed in different experiments (Chinnusamy et al., 2003; Zarka et al., 2003; Vogel et al., 2005; Agarwal et al., 2006; Medina et al., 2011); (2) an amplitude in the range of about 20–500 fold-increase compared to the steady state level prior to the stress (Chinnusamy et al., 2003; Zarka et al., 2003; Dong et al., 2011); and (3) a range of 3–6 h for the half-width of the peak in *CBF3* expression induced by the cold signal. For the brief initial Ca^{2+} pulse, we chose values for the kinetic parameters ensuring that the pulse is over in minutes, as observed in the experiments (Plieth et al., 1999). For Ca^{2+} oscillations, we selected parameter values in eqs. (2a)–(2d) as described in Section 1.2 of **Supplementary Material**, so as to obtain a period of the order of 150 s as observed in plant guard cells (Allen et al., 2001). The degradation rate of *CBF3* mRNA has been determined experimentally. Thus, the *CBF3* transcript has a half-life of 7.5 min at warm temperatures, which gives an approximately degradation rate of *CBF3* mRNA at 20°C of the order of 5.5 nM/h, while at low temperature this degradation rate is 10 times smaller, of the order of 0.55 nM/h (Zarka et al., 2003). This quantitative information was taken into account in numerical simulations of the model.

We also performed a sensitivity analysis, as detailed in Section 4.2 in **Supplementary Material**, to determine for each parameter, varying one at a time, the range in which the model matches experimental observations with respect to three properties of the peak of *CBF3* expression elicited by a cold stress: the amplitude of the peak in *CBF3* mRNA, the timing of this peak, and the half-width, measuring the duration of the response. This analysis allows us to determine which parameters might be targeted experimentally to modify most effectively the amplitude, timing, or half-width of the peak in *CBF3* expression elicited by the cold stress.

An additional interest of the computational model is to reveal the possibility of new modes of behavior, such as oscillations in *CBF3* expression, which, so far, have not been observed experimentally, and to investigate the effect of coupling the plant cold response pathway to the circadian clock. In a certain range of parameter values, the model indeed predicts the possible occurrence of sustained oscillations in the expression of *CBF3*. The phenomenon is accompanied by oscillations in ZAT12 but not in the other state variables of the network. The mechanism of CBF3 oscillations involves the negative feedback exerted by ZAT12 on the expression of *CBF3*. Numerical simulations indicate, as shown in "Putative CBF3 Oscillations Resulting From Negative Feedback Exerted by ZAT12" section and **Supplementary Figure S4**, that oscillations are favored by a high level of phosphorylated ICE1, which acts as inducer of *CBF3*. Besides augmenting the stability of *CBF3* mRNA, enhancing the level of phosphorylated ICE1 might, therefore, be a promising approach for investigating the possibility of an oscillatory expression of the genes that control the plant response to cold stress. The possible physiological significance of such oscillations in *CBF3* expression and their very occurrence nevertheless remain an open question, given that they only occur for appropriate parameter values and that so far they have not been observed experimentally. Besides these ZAT12-controlled oscillations, the cold response pathway is subjected to periodic alternations between cold and warm periods and to control by the endogenous plant circadian clock, which appears to gate the plant response to cold stress, as shown in "Circadian Gating" section and **Figure 8**. Because of their tight coupling through CBF3, ZAT12-controlled oscillations, if they occur, might well synchronize with these endogenous and exogenous circadian variations.

Incorporation of the effect of the circadian clock on *CBF3* expression into the model allowed us to account for circadian gating of the response to a drop in temperature. We represented the circadian input phenomenologically by means of a sine function multiplying the rate of *CBF3* mRNA synthesis. It would be possible to replace this description by using a compact model for the plant circadian clock (De Caluwé et al., 2016) in which CCA1 and LHY are treated explicitly as variables. This would allow us to test more directly the effect of different photoperiods as well as the effect of mutations in clock genes on the plant response to cold stress. In support of experimental observations (Fowler et al., 2005; Dong et al., 2011), the results of **Figure 8** indicate that the operation of the cold response pathway is gated by the circadian clock. Experimentally, the maximum expression of *CBF3* occurs during the day, several hours before the L to D transition. This timing does not correspond to the time at which the cold stress is expected to occur in physiological conditions, which should be in late evening or at night. The same remark holds for the results of **Figure 8**, which indicate that in the model also the window of maximum responsiveness of the cold response pathway is largely confined to the L phase. The results predict that in a portion of the LD cycle, the reaction of the plant cold response pathway to a Ca^{2+} pulse is negligible. From a physiological point of view, we would not expect the cold response pathway to operate optimally if the cold stress arrives at such phases. In discussing the relative timing of the system's response with respect to the cold stress, we should keep in mind that a delay likely exists between the rise in *CBF3* expression and the build-up of the cellular response to cold stress. A large number of *COR* genes must indeed be induced by CBF3, and their translation into the corresponding proteins takes time, while additional delays should arise from the time taken by these proteins to organize cold acclimation and freezing tolerance at the cellular level.

Besides the pathway responsible for resistance to a cold stress, plants display a separate cold-dependent response, which occurs on a time scale of weeks or months. This major cold-dependent physiological process, known as vernalization, is the phenomenon by which prolonged cold exposure leads to the silencing of a floral repressor gene (Song et al., 2012), thereby allowing the plant to delay its flowering until after the winter season. This epigenetic process has been modeled mathematically in terms of bistability in the flower-control regulatory system (Angel et al., 2011). Here, we focused on modeling the rapid response of plants to a cold stress that leads, after a few hours, to freezing tolerance. Although they appear to be largely independent and occur on widely different time scales, some experimental observations provide evidence (Diallo et al., 2010; Feng et al., 2017) for cross-talk between the two cold-dependent signaling pathways.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

XY, XH, and AG designed the study. RZ, AG, and DG built the computational model. RZ and DG performed numerical simulations. RZ and AG wrote the manuscript. RZ, DG, XH, XY, and AG contributed to discussing the results and to finalizing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

XY was supported by National Natural Science Foundation of China (no. 11171155, no. 11871268) and Natural Science Foundation of Jiangsu Province, China (no. BK20171370). XH was supported by the Key Projects of National Key Research and Development Plan (2017YFD0101803), the State Key Program of Natural Science Foundation of China (31330067), and the China Agriculture Research System (CARS-23-A-06). RZ was supported by the scholarship from China Scholarship Council (file no. 201706850036) during his stay at the Université Libre de Bruxelles (Brussels, Belgium). DG acknowledges support from the Fonds de la Recherche Scientifique (F.R.S.-FNRS, Belgium), CDR no J.0076.19.

ACKNOWLEDGMENTS

We wish to thank the referees for fruitful suggestions and Chunzhao Zhao for a helpful discussion. Part of this work was performed during the participation of RZ in the program "Certificat international de formation à la recherche doctorale" at the Université Libre de Bruxelles (ULB, Brussels, Belgium).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.591073/full#supplementary-material

REFERENCES

Achard, P., Gong, F., Cheminant, S., Alloua, M., Hedden, P., and Genschik, P. (2008). The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism. *Plant Cell* 20, 2117–2119. doi: 10.1105/tpc.108.058941

Agarwal, M., Hao, Y., Kapoor, A., Dong, C., Fujii, H., Zheng, X., et al. (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* 281, 37636–37645. doi: 10.1074/jbc.M605895200

Allen, G. J., Chu, S. P., Harrington, C. L., Schumacher, K., Hoffmann, T., Tang, Y. Y., et al. (2001). A defined range of guard cell calcium oscillation parameters encodes stomatal movements. *Nature* 411, 1053–1057. doi: 10.1038/35082575

Allen, G. J., Chu, S. P., Schumacher, K., Shimazaki, C. T., Vafeados, D., Kemper, A., et al. (2000). Alteration of stimulus-specific guard cell calcium oscillations and stomatal closing in *Arabidopsis* det3 mutant. *Science* 289, 2338–2342. doi: 10.1126/science.289.5488.2338

Angel, A., Song, J., Dean, C., and Howard, M. (2011). A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* 476, 105–108. doi: 10.1038/nature10241

Barah, P., Jayavelu, N. D., Rasmussen, S., Nielsen, H. B., Mundy, J., and Bones, A. (2013). Genome-scale cold stress response regulatory networks in ten *Arabidopsis thaliana* ecotypes. *BMC Genomics* 14, 722–737. doi: 10.1186/1471-2164-14-722

Bieniawska, Z., Espinoza, C., Schlereth, A., Sulpice, R., Hincha, D. K., and Hannah, M. A. (2008). Disruption of the *Arabidopsis* circadian clock is responsible for extensive variation in the cold-responsive transcriptome. *Plant Physiol.* 147, 263–279. doi: 10.1104/pp.108.118059

Chew, Y. H., and Halliday, K. J. (2010). A stress-free walk from *Arabidopsis* to crops. *Curr. Opin. Biotechnol.* 22, 1–6. doi: 10.1016/j.copbio.2010.11.011

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B., Hong, X., Agarwal, M., et al. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes Dev.* 17, 1043–1054. doi: 10.1101/gad.1077503

Chinnusamy, V., and Zhu, J. K. (2002). Molecular genetic analysis of cold-regulated gene transcription. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 357, 877–886. doi: 10.1098/rstb.2002.1076

Chinnusamy, V., Zhu, J. H., and Zhu, J. K. (2006). Gene regulation during cold acclimation in plants. *Physiol. Plant.* 126, 52–61. doi: 10.1111/j.1399-3054.2006.00596.x

Chinnusamy, V., Zhu, J. H., and Zhu, J. K. (2007). Cold stress regulation of gene expression in plants. *Trends Plant Sci.* 12, 444–451. doi: 10.1016/j.tplants.2007.07.002

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M. F. (2004). A prominent role for the CBF cold responsive pathway in configuring the low-temperature metabolome of *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15243–15248. doi: 10.1073/pnas.0406069101

De Caluwé, J., Xiao, Q., Hermans, C., Verbruggen, N., Leloup, J. -C., and Gonze, D. (2016). A compact model for the complex plant circadian clock. *Front. Plant Sci.* 7:74. doi: 10.3389/fpls.2016.00074

Diallo, A., Kane, N., Agharbaoui, Z., Badawi, M., and Sarhan, F. (2010). Heterologous expression of wheat VERNALIZATION 2 (TaVRN2) gene in *Arabidopsis* delays flowering and enhances freezing tolerance. *PLoS One* 5:e8690. doi: 10.1371/journal.pone.0008690

Ding, Y., Shi, Y., and Yang, S. (2019). Advances and challenges in uncovering cold tolerance regulatory mechanisms in plants. *New Phytol.* 222, 1690–1704. doi: 10.1111/nph.15696

Dodd, A. N., Jakobsen, M. K., Baker, A. J., Telzerow, A., Hou, S., Laplaze, L., et al. (2006). Time of day modulates low-temperature Ca^{2+} signals in *Arabidopsis*. *Plant J.* 48, 962–973. doi: 10.1111/j.1365-313X.2006.02933.x

Doherty, C. J., Van Buskirk, H. A., Myers, S. J., and Thomashow, M. F. (2009). Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. *Plant Cell* 21, 972–984. doi: 10.1105/tpc.108.063958

Dong, C., Agarwal, M., Zhang, Y., Xie, Q., and Zhu, J. K. (2006). The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8281–8286. doi: 10.1073/pnas.0602874103

Dong, M. A., Farré, E. M., and Thomashow, M. F. (2011). Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7241–7246. doi: 10.1073/pnas.1103741108

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33, 751–763. doi: 10.1046/j.1365-313X.2003.01661.x

Dupont, G., Berridge, M. J., and Goldbeter, A. (1991). Signal-induced Ca^{2+} oscillations: properties of a model based on Ca^{2+}-induced Ca^{2+} release. *Cell Calcium* 12, 73–85. doi: 10.1016/0143-4160(91)90010-c

Ermentrout, B. (2002). *Simulating, analyzing, and animating dynamical systems: A guide to XPPAUT for researchers and students*. Philadelphia: SIAM.

Feng, Y., Yin, Y., and Fei, S. (2017). BdVRN1 expression confers flowering competency and is negatively correlated with freezing tolerance in *Brachypodium distachyon*. *Front. Plant Sci.* 8:1107. doi: 10.3389/fpls.2017.01107

Fowler, S. G., Cook, D., and Thomashow, M. F. (2005). Low temperature induction of *Arabidopsis* CBF1, 2, and 3 is gated by the circadian clock. *Plant Physiol.* 137, 961–968. doi: 10.1104/pp.104.058354

Fowler, S., and Thomashow, M. F. (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* 14, 1675–1690. doi: 10.1105/tpc.003483

Furuya, T., Matsuoka, D., and Nanmori, T. (2013). Phosphorylation of *Arabidopsis thaliana* MEKK1 via Ca^{2+} signaling as a part of the cold stress response. *J. Plant Res.* 126, 833–840. doi: 10.1007/s10265-013-0576-0

Gérard, C., and Goldbeter, A. (2009). Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21643–21648. doi: 10.1073/pnas.0903827106

Gilmour, S. J., Fowler, S. G., and Thomashow, M. F. (2004). *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. *Plant Mol. Biol.* 54, 767–781. doi: 10.1023/B:PLAN.0000040902.06881.d4

Goldbeter, A. (1995). A model for circadian oscillations in the *Drosophila* period protein (PER). *Proc. R. Soc. Lond. B Biol. Sci.* 261, 319–324. doi: 10.1098/rspb.1995.0153

Goldbeter, A. (1996). *Biochemical oscillations and cellular rhythms: The molecular bases of periodic and chaotic behavior*. Cambridge, UK: Cambridge University Press.

Goldbeter, A., Dupont, G., and Berridge, M. J. (1990). Minimal model for signal-induced Ca^{2+} oscillations and for their frequency encoding through protein phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* 87, 1461–1465. doi: 10.1073/pnas.87.4.1461

Goldbeter, A., Gérard, C., Gonze, D., Leloup, J. -C., and Dupont, G. (2012). Systems biology of cellular rhythms. *FEBS Lett.* 586, 2955–2965. doi: 10.1016/j.febslet.2012.07.041

Hannah, M. A., Heyer, A. G., and Hincha, D. K. (2005). A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. *PLoS Genet.* 1:e26. doi: 10.1371/journal.pgen.0010026

Hannah, M. A., Wiese, D., Freund, S., Fiehn, O., Heyer, A. G., and Hincha, D. K. (2006). Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiol.* 142, 98–112. doi: 10.1104/pp.106.081141

Hua, J. (2009). From freezing to scorching, transcriptional responses to temperature variations in plants. *Curr. Opin. Plant Biol.* 12, 568–573. doi: 10.1016/j.pbi.2009.07.012

Ishitani, M., Xiong, L., Lee, H., Stevenson, B., and Zhu, J. K. (1998). HOS1, a genetic locus involved in cold-responsive gene expression in *Arabidopsis*. *Plant Cell* 10, 1151–1161. doi: 10.1105/tpc.10.7.1151

Jaglo, K. R., Kleff, S., Amundsen, K. L., Zhang, X., Haake, V., Zhang, J. Z., et al. (2001). Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol.* 127, 910–917. doi: 10.1104/pp.010548

Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O., and Thomashow, M. F. (1998). *Arabidopsis* CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* 280, 104–106. doi: 10.1126/science.280.5360.104

Jiang, B., Shi, Y., Peng, Y., Jia, Y., Yan, Y., Dong, X., et al. (2020). Cold-induced CBF3-PIF3 interaction enhances freezing tolerance by stabilizing the phyB themosensor in *Arabidopsis*. *Mol. Plant* 13, 894–906. doi: 10.1016/j.molp.2020.04.006

Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., et al. (2004). Exploring the temperature-stress metabolome of *Arabidopsis*. *Plant Physiol.* 136, 4159–4168. doi: 10.1104/pp.104.052142

Keily, J., MacGregor, D. R., Smith, R. W., Millar, A. J., Halliday, K. J., and Penfield, S. (2013). Model selection reveals control of cold signaling by evening-phased components of the plant circadian clock. *Plant J.* 76, 247–257. doi: 10.1111/tpj.12303

Kim, S. H., Kim, H. S., Bahk, S., An, J., Yoo, Y., Kim, J. Y., et al. (2017). Phosphorylation of the transcriptional repressor MYB15 by mitogen-activated protein kinase 6 is required for freezing tolerance in *Arabidopsis*. *Nucleic Acids Res.* 45, 6613–6627. doi: 10.1093/nar/gkx417

Knight, H., Trewavas, A. J., and Knight, M. R. (1996). Cold calcium signaling in *Arabidopsis* involves two cellular pools and a change in calcium signature after acclimation. *Plant Cell* 8, 489–503. doi: 10.1105/tpc.8.3.489

Kong, Q., Qu, N., Gao, M., Zhang, Z., Ding, X., Yang, F., et al. (2012). The MEKK1-MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-activated protein kinase kinase kinase in *Arabidopsis*. *Plant Cell* 24, 2225–2236. doi: 10.1105/tpc.112.097253

Koshland, D. E., Goldbeter, A., and Stock, J. B. (1982). Amplification and adaption in regulatory and sensory systems. *Science* 217, 220–225. doi: 10.1126/science.7089556

Kudla, J., Becker, D., Grill, E., Hedrich, R., Hippler, M., Kummer, U., et al. (2018). Advances and current challenges in calcium signaling. *New Phytol.* 218, 414–431. doi: 10.1111/nph.14966

Lee, B., Henderson, D. A., and Zhu, J. K. (2005). The *Arabidopsis* cold-responsive transcriptome and its regulation by ICE1. *Plant Cell* 17, 3155–3175. doi: 10.1105/tpc.105.035568

Lee, H., Xiong, L., Gong, Z., Ishitani, M., Stevenson, B., and Zhu, J. K. (2001). The *Arabidopsis* HOS1 gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo–cytoplasmic partitioning. *Genes Dev.* 15, 912–924. doi: 10.1101/gad.866801

Leloup, J. -C., and Goldbeter, A. (2003). Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7051–7056. doi: 10.1073/pnas.1132112100

Leloup, J. -C., and Goldbeter, A. (2004). Modeling the mammalian circadian clock: sensitivity analysis and multiplicity of oscillatory mechanisms. *J. Theor. Biol.* 230, 541–562. doi: 10.1016/j.jtbi.2004.04.040

Li, H., Ding, Y., Shi, Y., Zhang, X., Zhang, S., Gong, Z., et al. (2017). MPK3- and MPK6-mediated ICE1 phosphorylation negatively regulates ICE1 stability and freezing tolerance in *Arabidopsis*. *Dev. Cell* 43, 1–13. doi: 10.1016/j.devcel.2017.09.025

Lissarre, M., Ohta, M., Sato, A., and Miura, K. (2010). Cold-responsive gene regulation during cold acclimation in plants. *Plant Signal. Behav.* 5, 948–952. doi: 10.4161/psb.5.8.12135

Liu, Y., Dang, P., Liu, L., and He, C. (2019). Cold acclimation by the CBF-COR pathway in a changing climate: lessons from *Arabidopsis thaliana*. *Plant Cell Rep.* 38, 511–519. doi: 10.1007/s00299-019-02376-3

Liu, Z., Jia, Y., Ding, Y., Shi, Y., Li, Z., Guo, Y., et al. (2017). Plasma membrane CRPK1-mediated phosphorylation of 14-3-3 proteins induces their nuclear import to fine-tune CBF signaling during cold response. *Mol. Cell* 66, 117.e5–128.e5. doi: 10.1016/j.molcel.2017.02.016

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., et al. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain, separate two cellular signal transduction pathways in drought- and low temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* 10, 1391–1406. doi: 10.1105/tpc.10.8.1391

Marozsán-Tóth, Z., Vashegyi, I., Galiba, G., and Tóth, B. (2015). The cold response of CBF genes in barley is regulated by distinct signaling mechasnims. *J. Plant Physiol.* 181, 42–49. doi: 10.1016/j.jplph.2015.04.002

McClung, C. R. (2019). The plant circadian oscillator. *Biology* 8:14. doi: 10.3390/biology8010014

McClung, C. R., and Davis, S. J. (2010). Ambient thermometers in plants: from physiological outputs towards mechanisms of thermal sensing. *Curr. Biol.* 20, R1086–R1092. doi: 10.1016/j.cub.2010.10.035

Medina, J., Catalá, R., and Salinas, J. (2011). The CBFs: three *Arabidopsis* transcription factors to cold acclimate. *Plant Sci.* 180, 3–11. doi: 10.1016/j.plantsci.2010.06.019

Miura, K., Jin, J. B., Lee, J., Yoo, C. Y., Stirm, V., Miura, T., et al. (2007). SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in *Arabidopsis*. *Plant Cell* 19, 1403–1414. doi: 10.1105/tpc.106.048397

Nakamichi, N., Kusano, M., Fukushima, A., Kita, M., Ito, S., Yamashino, T., et al. (2009). Transcript profiling of an *Arabidopsis* PSEUDO RESPONSE REGULATOR arrhythmic triple mutant reveals a role for the circadian clock in cold stress response. *Plant Cell Physiol.* 50, 447–462. doi: 10.1093/pcp/pcp004

Nohales, M. A., and Kay, S. A. (2016). Molecular mechanisms at the core of the plant circadian oscillator. *Nat. Struct. Mol. Biol.* 23, 1061–1069. doi: 10.1038/nsmb.3327

Örvar, B. L., Sangwan, V., Omann, F., and Dhindsa, R. S. (2000). Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity. *Plant J.* 23, 785–794. doi: 10.1046/j.1365-313x.2000.00845.x

Plieth, C. (1999). Temperature sensing by plants: calcium-permeable channels as primary sensors—a model. *J. Membr. Biol.* 172, 121–127. doi: 10.1007/s002329900590

Plieth, C., Hansen, U., Knight, H., and Knight, M. R. (1999). Temperature sensing by plants: the primary characteristics of signal perception and calcium response. *Plant J.* 18, 491–497. doi: 10.1046/j.1365-313x.1999.00471.x

Ruelland, E., and Zachowski, A. (2010). How plants sense temperature. *Environ. Exp. Bot.* 69, 225–232. doi: 10.1016/j.envexpbot.2010.05.011

Ruiz, M. C. M., Hubbard, K. E., Gardner, M. J., Jung, H. J., Aubry, S., Hotta, C. T., et al. (2018). Circadian oscillations of cytosolic free calcium regulate the *Arabidopsis* circadian clock. *Nat. Plants* 4, 690–698. doi: 10.1038/s41477-018-0224-8

Ruiz, M. C. M., Jung, H. J., and Webb, A. A. R. (2020). Circadian gating of dark-induced increases in chloroplast- and cytosolic-free calcium in *Arabidopsis*. *New Phytol.* 225, 1993–2005. doi: 10.1111/nph.16280

Shi, Y., Ding, Y., and Yang, S. (2018). Molecular regulation of CBF signaling in cold acclimation. *Trends Plant Sci.* 23, 623–637. doi: 10.1016/j.tplants.2018.04.002

Song, J., Angel, A., Howard, M., and Dean, C. (2012). Vernalization—a cold-induced epigenetic switch. *J. Cell Sci.* 125, 3723–3731. doi: 10.1242/jcs.084764

Stockinger, E. J., Gilmour, S. J., and Thomashow, M. F. (1997). *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcription activator that binds to the C repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc. Natl. Acad. Sci. U. S. A.* 94, 1035–1040. doi: 10.1073/pnas.94.3.1035

Tang, K., Zhao, L., Ren, Y., Yang, S., Zhu, J., and Zhao, C. (2020). The transcription factor ICE1 functions in cold stress response by binding to the promoters of *CBF* and *COR* genes. *J. Integr. Plant Biol.* 62, 258–263. doi: 10.1111/jipb.12918

Teige, M., Scheikl, E., Eulgem, T., Dóczi, R., Ichimura, K., Shinozaki, K., et al. (2004). The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol. Cell* 15, 141–152. doi: 10.1016/j.molcel.2004.06.023

Thomashow, M. F. (1999). Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 571–599. doi: 10.1146/annurev.arplant.50.1.571

Thomashow, M. F. (2010). Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiol.* 154, 571–577. doi: 10.1104/pp.110.161794

Vogel, J. T., Zarka, D. G., Van Buskirk, H. A., Fowler, S. G., and Thomashow, M. F. (2005). Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. *Plant J.* 41, 195–211. doi: 10.1111/j.1365-313X.2004.02288.x

Wang, Z., Wang, F., Tang, J., Huang, Z., Xiong, A., and Hou, X. (2014). C-repeat binding factor gene family identified in non-heading Chinese cabbage is functional in abiotic and biotic stress response but different from that in *Arabidopsis*. *Acta Physiol. Plant.* 36, 3217–3229. doi: 10.1007/s11738-014-1688-4

Yang, T., Ali, G. S., Yang, L., Du, L., Reddy, A. S. N., and Poovaiah, B. W. (2010). Calcium/calmodulin-regulated receptor-like kinase CRLK1 interacts with MEKK1 in plants. *Plant Signal. Behav.* 5, 991–994. doi: 10.4161/psb.5.8.12225

Zarka, D. G., Vogel, J. T., Cook, D., and Thomashow, M. F. (2003). Cold induction of *Arabidopsis* CBF genes involves multiple ICE (inducer of CBF expression) promoter elements and a cold-regulatory circuit that is desensitized by low temperature. *Plant Physiol.* 133, 910–918. doi: 10.1104/pp.103.027169is

Zhang, X., Fowler, S. G., Cheng, H., Lou, Y., Rhee, S. Y., Stockinger, E. J., et al. (2004). Freezing-sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing-tolerant *Arabidopsis*. *Plant J.* 39, 905–919. doi: 10.1111/j.1365-313X.2004.02176.x

Zhao, C., Wang, P., Si, T., Hsu, C. C., Wang, L., Zayed, O., et al. (2017). MAP kinase cascades regulate the cold response by modulating ICE1 protein stability. *Dev. Cell* 43, 1–12. doi: 10.1016/j.devcel.2017.09.024

Zhao, C., Zhang, Z., Xie, S., Si, T., Li, Y., and Zhu, J. -K. (2016). Mutational evidence for the critical role of CBF transcription factors in cold acclimation in *Arabidopsis*. *Plant Physiol.* 171, 2744–2759. doi: 10.1104/pp.16.00533

Zielinski, R. E. (1998). Calmodulin and calmodulin-binding proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 697–725. doi: 10.1146/annurev.arplant.49.1.697

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 12 November 2020
doi: 10.3389/fphys.2020.590479

Dynamical Boolean Modeling of Immunogenic Cell Death

Andrea Checcoli[1], Jonathan G. Pol[2,3], Aurelien Naldi[1], Vincent Noel[4,5,6], Emmanuel Barillot[4,5,6], Guido Kroemer[2,3,7,8,9], Denis Thieffry[1], Laurence Calzone[4,5,6†] and Gautier Stoll[2,3*†]*

[1] Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France, [2] Equipe 11 labellisée par la Ligue Nationale contre le Cancer, Centre de Recherche des Cordeliers, INSERM U1138, Université de Paris, Sorbonne Université, Institut Universitaire de France, Paris, France, [3] Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus, Villejuif, France, [4] Institut Curie, PSL Research University, Paris, France, [5] INSERM, U900, Paris, France, [6] MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France, [7] Pôle de Biologie, Hôpital Europeen Georges Pompidou, AP-HP, Paris, France, [8] Suzhou Institute for Systems Biology, Chinese Academy of Sciences, Suzhou, China, [9] Department of Women's and Children's Health, Karolinska University Hospital, Stockholm, Sweden

As opposed to the standard tolerogenic apoptosis, immunogenic cell death (ICD) constitutes a type of cellular demise that elicits an adaptive immune response. ICD has been characterized in malignant cells following cytotoxic interventions, such as chemotherapy or radiotherapy. Briefly, ICD of cancer cells releases some stress/danger signals that attract and activate dendritic cells (DCs). The latter can then engulf and cross-present tumor antigens to T lymphocytes, thus priming a cancer-specific immunity. This series of reactions works as a positive feedback loop where the antitumor immunity further improves the therapeutic efficacy by targeting cancer cells spared by the cytotoxic agent. However, not all chemotherapeutic drugs currently approved for cancer treatment are able to stimulate bona fide ICD: some commonly used agents, such as cisplatin or 5-fluorouracil, are unable to activate all features of ICD. Therefore, a better characterization of the process could help identify some gene or protein candidates to target pharmacologically and suggest combinations of drugs that would favor/increase antitumor immune response. To this end, we have built a mathematical model of the major cell types that intervene in ICD, namely cancer cells, DCs, $CD8^+$ and $CD4^+$ T cells. Our model not only integrates intracellular mechanisms within each individual cell entity, but also incorporates intercellular communications between them. The resulting cell population model recapitulates key features of the dynamics of ICD after an initial treatment, in particular the time-dependent size of the different cell types. The model is based on a discrete Boolean formalism and is simulated by means of a software tool, UPMaBoSS, which performs stochastic simulations with continuous time, considering the dynamics of the system at the cell population level with appropriate timing of events, and accounting for death and division of each cell type. With this model, the time scales of some of the processes involved in ICD, which are challenging to measure experimentally, have been predicted. In addition, our model analysis led to the identification of actionable targets for boosting ICD-induced antitumor response. All computational analyses and results are compiled in interactive notebooks which cover the presentation of the network structure, model simulations, and parameter sensitivity analyses.

Keywords: logical modeling, immunogenic cell death, antitumor immune response, dendritic cells, cytotoxic $CD8^+$ T lymphocytes

Edited by:
Jianhua Xing,
University of Pittsburgh, United States

Reviewed by:
Tongli Zhang,
University of Cincinnati, United States
Huijing Wang,
University of Pittsburgh, United States

****Correspondence:***
Laurence Calzone
laurence.calzone@curie.fr
Gautier Stoll
gautier.stoll@upmc.fr

† These authors share senior authorship

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *01 August 2020*
Accepted: *24 October 2020*
Published: *12 November 2020*

Citation:
Checcoli A, Pol JG, Naldi A, Noel V, Barillot E, Kroemer G, Thieffry D, Calzone L and Stoll G (2020) Dynamical Boolean Modeling of Immunogenic Cell Death. Front. Physiol. 11:590479. doi: 10.3389/fphys.2020.590479

1. INTRODUCTION

In order to ensure tissue homeostasis, programmed cell death modalities such as intrinsic apoptosis normally eliminates tumorigenic cells. When such initial safety mechanisms have failed, the immune system intervenes as a backup system to eliminate or control carcinogenic lesions (Galluzzi et al., 2017). Innate and adaptive immune actors, embodied by natural killer cells and cytotoxic CD8$^+$ T cells, can mediate either extrinsic apoptosis or lysis of the aberrant cellular targets (Kroemer et al., 2013; Galluzzi et al., 2017). However, under immune pressure, some malignant cells able to evade immunosurveillance can be selected, proliferate, and generate a tumor mass. Yet, some cancer treatments, including several chemotherapies, radiotherapy, or oncolytic virotherapy, have the ability to reinstate cancer immunosurveillance (Galluzzi et al., 2017; Kepp et al., 2018). These cytotoxic interventions trigger a particular demise of transformed cells called "immunogenic cell death" (ICD) (Galluzzi et al., 2017; Kepp et al., 2018). As opposed to standard apoptosis, which is tolerogenic, ICD refers to an apoptotic process that elicits an adaptive immune response against tumor cells (**Figure 1**).

Following administration of an ICD-inducing treatment, dying cancer cells expose or release damaged-associated molecular patterns (DAMPs) together with tumor antigens in the tumor microenvironment (Fucikova et al., 2015; Galluzzi et al., 2017). The presence of such stress signals in the extracellular milieu allows the recruitment and activation of antigen-presenting cells, such as dendritic cells (DCs) (Galluzzi et al., 2017; Giovanelli et al., 2019). Counting among the DAMPs, extracellular ATP can be sensed by DCs through the purinergic receptors P2RX7 and P2RY2, and triggers their migration to the tumor bed and their activation (Elliott et al., 2009; Saez et al., 2017). An additional stress signal recorded upon ICD consists of calreticulin (CALR), a highly conserved chaperone protein residing in the lumen of the endoplasmic reticulum (ER) (Panaretakis et al., 2009). Upon stress of the ER, CALR translocates at the surface of the plasma membrane where it is detected by immature DCs through their low-density lipoprotein receptor-related protein 1 (LRP1)

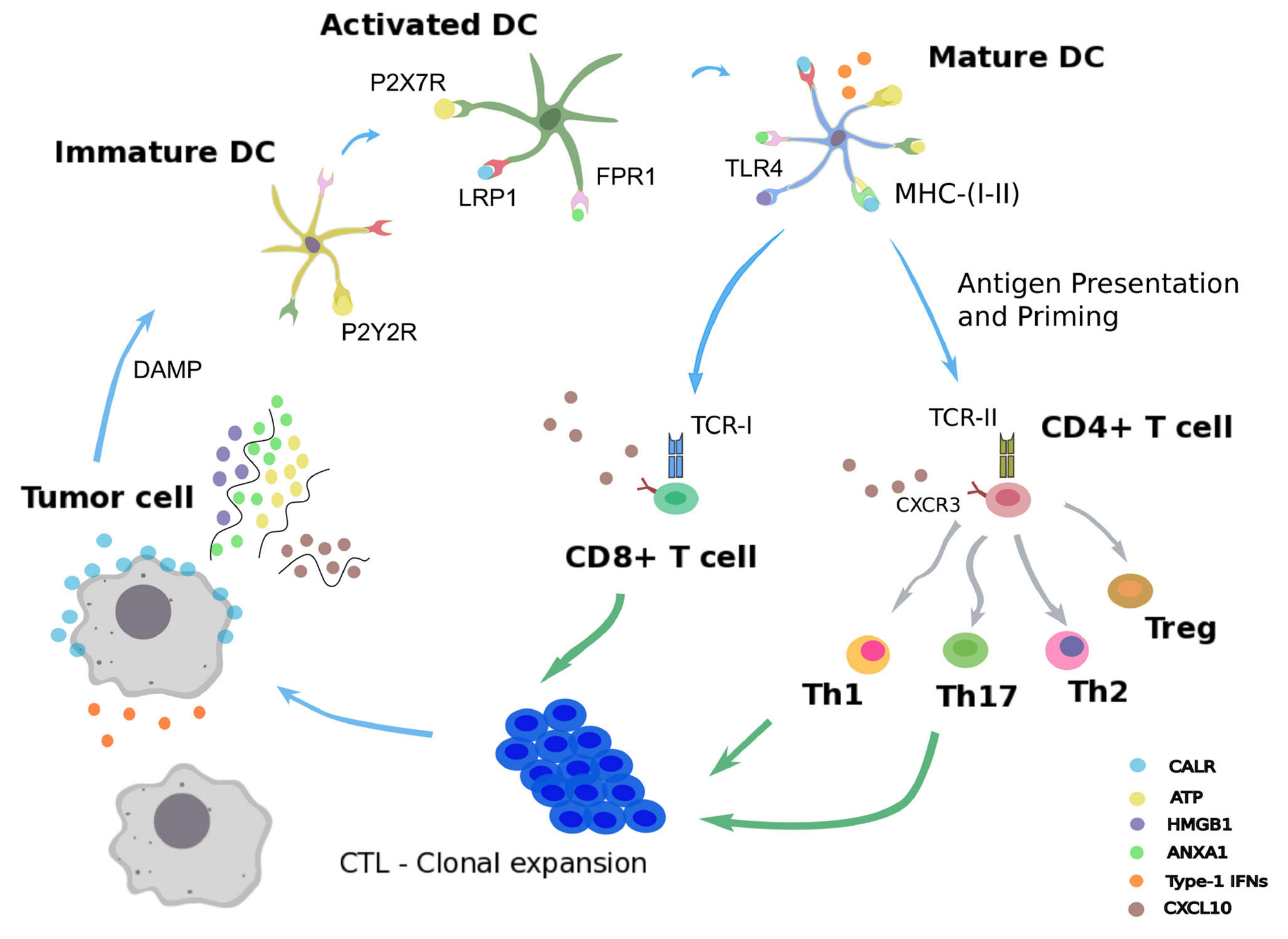


FIGURE 1 | Immunogenic cell death. Schematic representation of the immunogenic cell death cycle, starting with release of DAMPs from a dying tumor cell, leading to the maturation of a dendritic cell (DC), ultimately activating CD4$^+$ and CD8$^+$ T cells, which in turn trigger the death of the remaining live tumor cells.

(Gardai et al., 2005). Interaction of CALR with LRP1 acts as an "eat-me" signal, which promotes tumor antigen uptake by immature DCs (Galluzzi et al., 2017; Kepp et al., 2018). The high molecular group B1 protein (HMGB1) is the most abundant non-histone chromatin-binding protein. While restricted to the nucleus in normal condition, HMGB1 is freed in the tumor microenvironment upon ICD, thereby constituting another "alarm" signal for the immune system (Dumitriu et al., 2007). HMGB1 is detected by DCs through different receptors, mainly toll-like receptor 4 (TLR4), and promotes their activation (Apetoh et al., 2007). Similarly, cancer cells undergoing ICD release the cytoplasmic protein annexin A1 (ANXA1), another DAMP that binds to formyl peptide receptor 1 (FPR1) at the surface of DCs (Vacchelli et al., 2015; Galluzzi et al., 2017; Kepp et al., 2018). Local detection of ANXA1 by DCs contributes to their homing at proximity of dying malignant cells. Finally, ICD mimics a viral infection in cancer cells by triggering the production of type 1 interferons (IFN1). These latter act in an autocrine and paracrine manner to stimulate the secretion of the chemokine C-X-C motif chemokine ligand 10 (CXCL10) by cancer cells and thus favor the recruitment of activated T cells expressing its receptor, C-X-C motif chemokine receptor 3 (CXCR3) (Galluzzi et al., 2017; Kepp et al., 2018).

Once recruited to the tumor bed by ICD-related DAMPs, activated DCs can capture tumor antigens and undergo maturation. Thus, DCs upregulate the surface class-I and class-II major histocompatibility complex molecules (MHC-I, MHC-II) and cross-present tumor epitopes onto them (Galluzzi et al., 2017; Giovanelli et al., 2019). Additionally, DCs express co-stimulatory molecules, including CD40, CD80, and CD86, and secrete inflammatory cytokines, such as interleukin-12 (IL-12), IL-6, and tumor necrosis factor alpha (TNF-α) (Galluzzi et al., 2017; Giovanelli et al., 2019). In parallel, DCs upregulate the lymphoid tissue-residing C-C chemokine receptor type 7 (CCR7) that promote their migration to the draining lymph node (Riol-Blanco et al., 2005; Galluzzi et al., 2017). In the lymph node, mature DCs prime both naive CD4$^+$ and CD8$^+$ T lymphocytes that display the cognate T cell receptors (TCRs) (Galluzzi et al., 2017). Activated CD4$^+$ T cells can differentiate into conventional T helper cells (Th) or into regulatory T cells (Treg) (Li and Rudensky, 2016; Zhu, 2018). Depending on the cytokines locally present, several Th lineages can be distinguished, such as Th1 and Th17, which are involved in cancer immunosurveillance (Zhu, 2018). By producing type-1 cytokines such as interleukin-2 (IL-2) and IFN-γ, Th1 CD4$^+$ T cells actually support the differentiation of activated CD8$^+$ T cells (preCTL) into type 1 cytotoxic CD8$^+$ T lymphocytes (Tc1 / CTLs), which play a critical role in eliminating malignant entities (Kurokawa et al., 2001). Activated Th1 lymphocytes and CTLs are able to migrate from the lymph node to the blood stream and eventually reach the tumor site in a CXCL10-dependent manner (Galluzzi et al., 2017; Kepp et al., 2018). In the tumor site, Th1 cells can mediate antitumor activity via the secretion of the effector cytokine IFN-γ and the CTL-mediated release of the pore-forming perforins and cytotoxic granzymes (Kepp et al., 2018; Farhood et al., 2019). This series of reactions triggered by cancer ICD form a positive feedback loop where the mounted antitumor immunity further improves the therapeutic efficacy by targeting cancer cells spared by the cytotoxic agent (Galluzzi et al., 2017; Kepp et al., 2018; Farhood et al., 2019).

In this study, our aim was to develop a model integrating the molecular and cellular entities involved in cancer ICD, together with the subsequent immune cascade resulting into antitumor activity. A better characterization of the process could help identify actionable molecular components and thus suggest combinations of pharmacological compounds that would favor/increase anticancer immunity. In particular, not all chemotherapeutic drugs currently approved for the care of cancer are able to stimulate *bona fide* ICD (Galluzzi et al., 2017). Some commonly used agents, such as cisplatin or 5-fluorouracil, fail to activate some features of ICD (Bezu et al., 2015). Nevertheless, experimental complementation of cisplatin with cardiac glycosides resulted in *bona fide* ICD and translated into potent immunotherapeutic efficacy (Kepp et al., 2012; Menger et al., 2012). An *in silico* model could accelerate the identification of such combination regimens.

To this end, we have assembled a mathematical model covering the major cell types and biological processes intervening in ICD. In order to preserve the feasibility of the simulations, we properly restricted the number of interactions and processes. Tumor cells, dendritic cells, CD4$^+$ and CD8$^+$ T lymphocytes have been selected as the main players. Our model focuses on intercellular communications between the different cell types, considers current knowledge on the timing of events, and takes into account death and proliferation of tumor cells in diverse contexts.

The model is based on a discrete Boolean formalism and is simulated by means of a software tool, UPMaBoSS, which performs stochastic simulations with continuous time, considering the dynamics of the system at the cell population level (an extended presentation of UPMaBoSS is provided in Stoll et al., 2020). The Boolean description of the model entities is a strong approximation. In the present work, it fits the current biological knowledge, which is mainly qualitative. However, the grammar of MaBoSS allows to represent discrete levels of each model entity, corresponding to different concentration levels or different status of a protein (e.g., phosphorylated, or in complex) entailing different levels of activity. An example of multi-level dynamics of the protein p53 in the context of p53/Mdm2 interaction is provided at the following address: https://github.com/sysbio-curie/MaBoSS-env-2.0/tree/master/tutorial/MaBoSS-2.0.

The resulting cell population model recapitulates key features of the dynamics of ICD after an initial treatment, in particular the time-dependent size of the different types of cell populations. Furthermore, the time scales of some of the processes involved in ICD, which are challenging to measure experimentally, have been predicted by proper simulations. In addition, the analysis of our model led to the identification of potential target components to modulate in order to boost ICD-induced antitumor response.

2. MATERIALS AND METHODS

For modeling ICD, we use the stochastic Boolean simulation framework UPMaBoSS (Stoll et al., 2020). It is based on the previously defined MaBoSS grammar (Stoll et al., 2012), which was extended to simulate the dynamics of the cell populations. Details about the software are provided below.

More information about software accessibility, use cases, and jupyter notebooks including the code used for all the analyses presented here are available on a GitHub repository: https://github.com/sysbio-curie/ICD.

2.1. Cell Population Described as a Probabilistic Boolean System

The model is based on a regulatory network (directed graph), where nodes represent molecules, cell types and processes, whereas arcs represent positive or negative influences between these entities. In this application, some arrows can also represent transformation of cells. For instance, an immature DC can become mature (example in **Figure 2**, from DC to ActDC). In the Boolean description, the variables corresponding to these entities can take two values, 1 for active or present, and 0 for inactive or absent. We define a *network state* as a set (or a vector) of Boolean values associated with each of these entities. We apply a probabilistic framework to these network states. More precisely, a probability is associated with each network state, and the sum of probabilities over all possible network states is equal to 1. The interpretation in terms of population is straightforward: the number of cells in a given network state is equivalent to the network state probability multiplied by the overall size of the cell population.

2.2. Signaling Pathways Described as MaBoSS Models

MaBoSS (Stoll et al., 2012, 2017) is a software dedicated to the modeling of signaling pathways based on a Boolean probabilistic description. Given a Boolean model, MaBoSS produces time dependent probabilities of network states. In MaBoSS, two transition rates (activation/inhibition) are associated with each node of the network. These rates are written in a specific

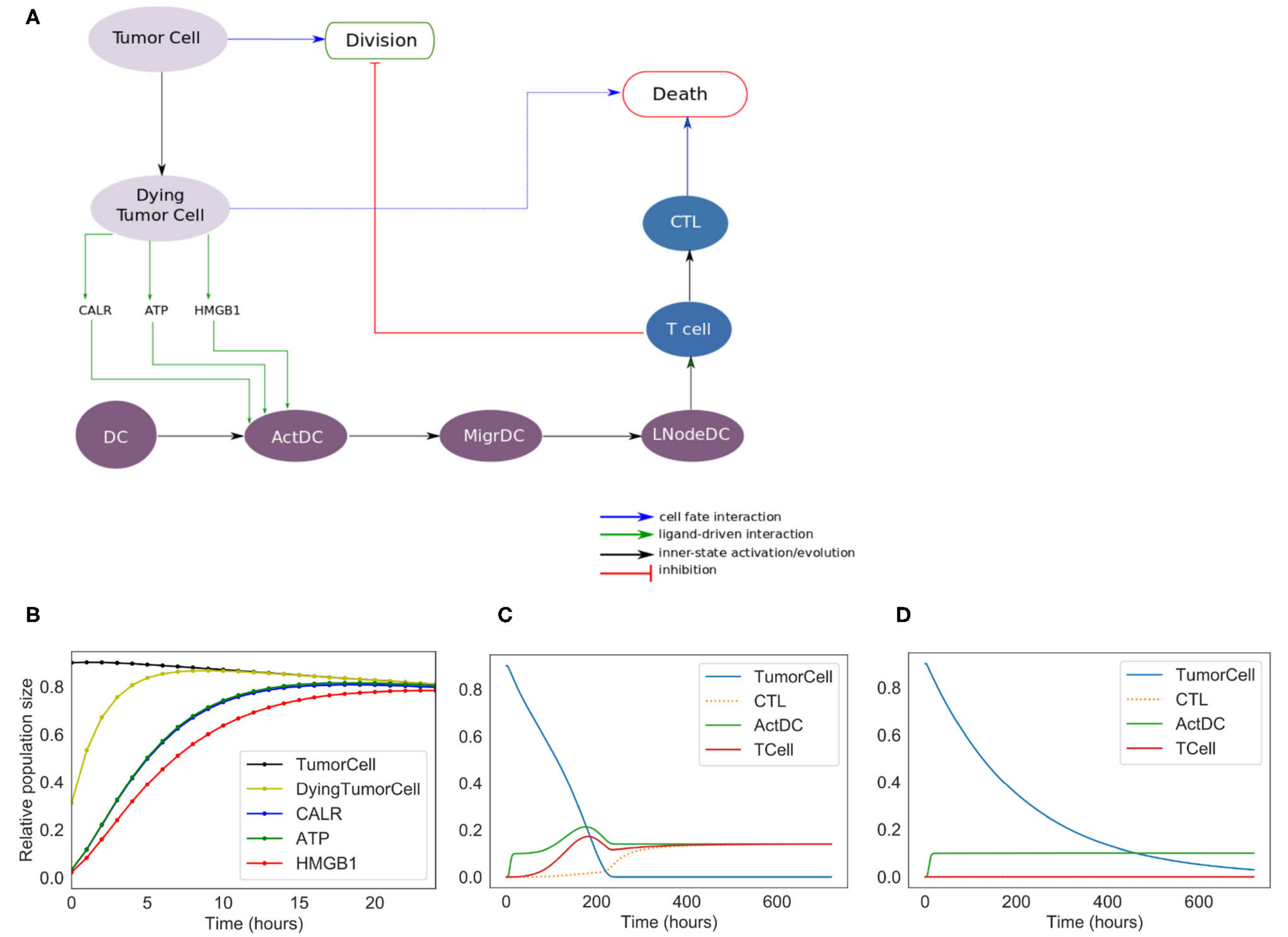


FIGURE 2 | Phenomenological model. **(A)** Influence network of the simplified version of ICD: it involves three cell types: tumor cells (Tumor Cell and Dying Tumor Cell for cells that have been treated by chemotherapeutic agents, gray nodes), dendritic cells (DC for immature dendritic cells, ActDC for mature dendritic cells, MigrDC for migrating dendritic cells, and LNodeDC for dendritic cells in the lymph node, purple nodes), T cells (T Cell, and CTL for cytotoxic T lymphocyte, blue nodes). **(B)** Early activation of ICD markers (Note: CALR overlaps with ATP). **(C)** Kinetics of the cell types with ICD. **(D)** Kinetics of the tumor cell populations without ICD.

grammar, which combines Boolean states of other nodes with logical (AND, OR, NOT) and arithmetic operators. In addition to this set of transition rates, a MaBoSS simulation requires the definition of an initial condition, i.e., a probability distribution of network states at time zero. In practice, a MaBoSS model is encoded into two files: a *bnd* file containing the logical rules and the definition of the transition rates, and a *cfg* file containing the parameters of the model (initial conditions and values for the transition rates) together with the simulation parameters.

2.3. Population Dynamics Described as an UPMaBoSS Model

MaBoSS can only model a cell population composed of non-interactive cells, with a constant population size. In order to better represent the dynamics of the cell populations, we recently developed an extension of MaBoSS, called UPMaBoSS (Stoll et al., 2020). In this framework, cells can divide, die and interact. In this respect, two additional nodes are added to the usual MaBoSS model: *Death* and *Division*. Regarding cell communication, ligand-receptor interactions are introduced: the transition rates of the receptors can contain parameters that are updated according to the state of the whole cell population. At regular time steps, the simulation is stopped and some variables are updated according to their status. Cells having the *Division* node at 1 are doubled, whereas those having the *Death* node at 1 are removed. Receptors are updated according to the value of their regulators.

An UPMaBoSS model is encoded into three files: *bnd* and *cfg* files written in MaBoSS grammar, completed by a specific *upp* file. The *upp* file contains the declaration of the *Death* and *Division* nodes, together with the updating rules for the parameters controlling the transition rates of the receptors.

All simulations of the ICD models reported here are encoded in the jupyter notebooks provided as **Supplementary Material**. These notebooks can be executed on any computer with a recent docker or a conda environment installed, as indicated in the dedicated GitHub repository (https://github.com/sysbio-curie/ICD). The full parameter sensitivity analysis is not included in the jupyter notebooks, as it is time consuming (taking over 24 h) and has been performed on a computer cluster. Results of the computations are available in **Supplementary Tables 4**, **5**. The notebooks contain some examples of models with modified parameters.

2.4. Phenomenological Model

The purpose of the ICD phenomenological model is to reproduce the succession of events observed experimentally, and to determine the role of each of the main cell types involved in ICD. This model serves as a basis to develop a more detailed, "extended" model. The phenomenological model includes three cell types with different status (**Figure 1**). *Tumor Cell* and *DC* constitute the inputs of the model. The simulation starts with a predefined tumor size and a given population of DCs. Tumor cells can die or divide. Dying tumor cells release some danger signals (CALR, ATP, and HMGB1), which activate DC (*ActDC*), which can then migrate (*MigrDC*) to the tumor draining Lymph Node (*LNodeDC*). There, mature DCs can activate T cells. T cells can then differentiate into cytotoxic effectors (*CTL*), proliferate, and reach the tumor through blood vessels. In contact with a CTL, tumor cells are cleared out. Without treatment or in the absence of T cells, tumor cells keep proliferating. The corresponding model files can be found at: https://github.com/sysbio-curie/ICD. The description of the model with the meaning of the variables and the parameter values are detailed in the **Supplementary Table 1**.

2.5. Extended Model

Some modeling choices were made to refine the phenomenological model, while preserving the overall dynamics. This extended model includes detailed representations of the series of events previously explored with the phenomenological model. The abstract transitions considered in the initial model are replaced by more refined details about which and how cells interact with ligand-receptor dynamics. Some of the ligands depend on the status of the cell type that produces them, and/or on the activity (or availability) of a receptor that mediates their activation. The transition rates associated with ligand activation are usually set as the inverse of the time that a ligand takes to reach its concentration peak.

In this extended model, we consider four cell types, including tumor cells, dendritic cells, CD4$^+$ T cells, and CD8$^+$ T cells, as shown in **Figure 3**. As for the phenomenological model, the size of the populations of these four cell types must be defined at the beginning of a simulation.

The extended model encompasses 57 entities, with some entities that may correspond to different instances of the same species or cell types (e.g., DC and aDC where "a" stands for active). **Figure 3** includes the cell types and their fate: *Tumor Cells* can die under treatment and become *Dying Tumor Cells*, dendritic cells (*DC*) can be activated in the presence of ATP, CD4$^+$ cells (*cd4*) can lead to Th1, Th2, Th17, or Treg cells, and finally CD8$^+$ cells (*cd8*) become *CTL*.

The cytokines and the corresponding receptors involved in ICD are each represented by a specific node (denoted in the model by *-o for a cytokine, and by *-rec for a receptor). The activation rule for each cytokine relies on the cell type producing it and possibly on the adjuvant effects of other ligands, enhancing the activation of such cytokines. On the other hand, a cytokine receptor node is conditioned by the presence of the cells presenting it. The two types of nodes, ligand and receptor, are dynamically linked, as the probability attributed to the ligand node impacts the rate of transition toward the activation of the receptor node. At each step of the simulation, after its first activation, the transition rate for the receptor is controlled by the ligands.

All the logical rules are written to account for the biological knowledge whenever it is reported in the literature. For instance, the rule for the CD8$^+$ T cell receptor node *TCR-I* is: `CD28 & CD8`, which reads as: when both CD28 (the receptor of CD80) and a naive CD8$^+$ T cell are present, TCR_I will be activated at a rate `$u_TCR_I$`, which is defined in the *cfg* file.

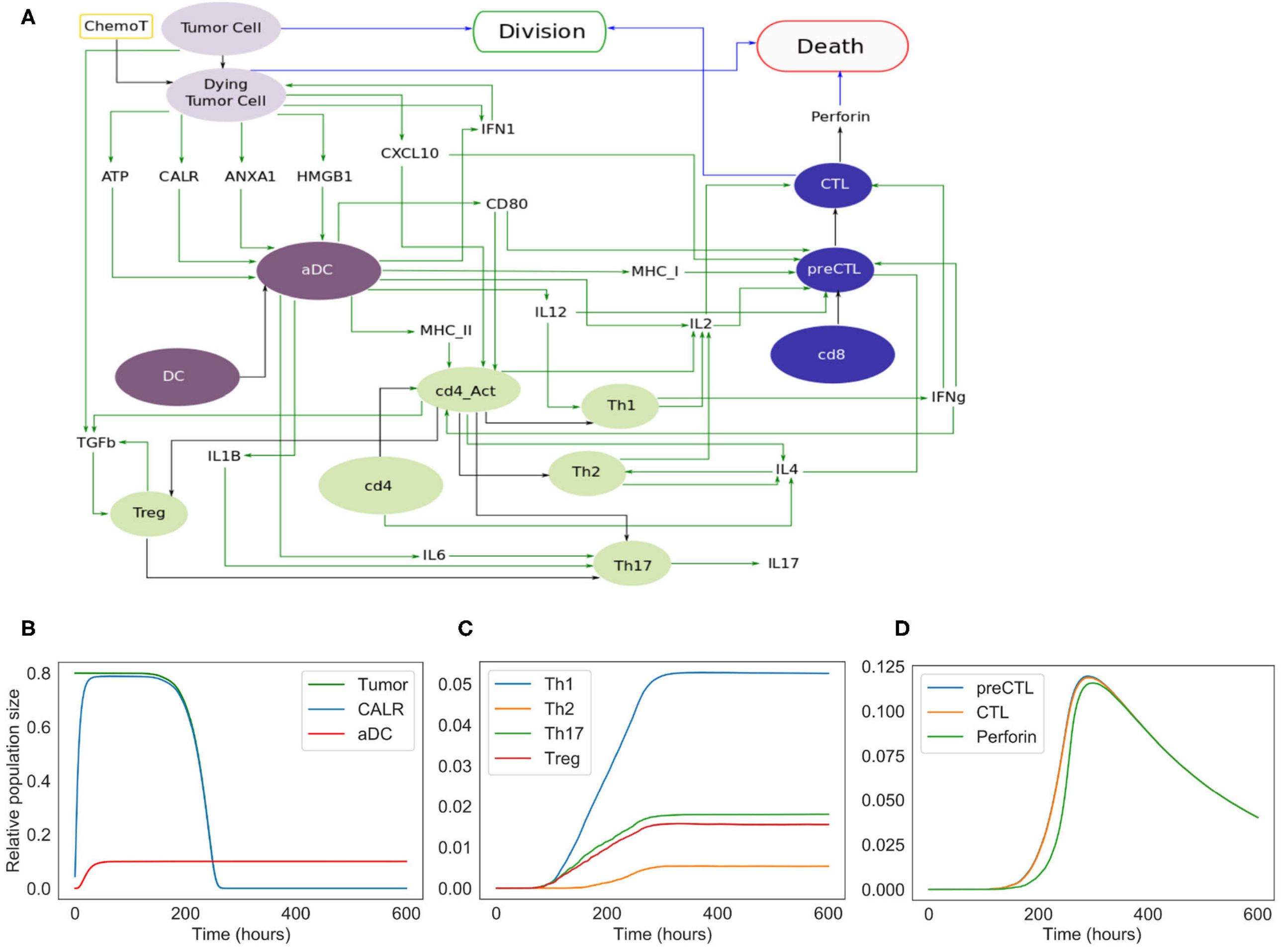


FIGURE 3 | Extended model. (A) Influence network of the extended version of ICD involving four cell types: tumor cells (light purple nodes), dendritic cells (dark purple nodes), CD4$^+$ T cells (green nodes) and CD8$^+$ T cells (blue nodes). Green arcs correspond to activation, black arcs to cell transformation, and blue arcs to cell fate interactions. Tumor Cell, DC, cd4, and cd8 are inputs. When the fifth input, namely ChemoT, is active, we consider that cells have incorporated a chemotherapeutic drug. The two cell fates, Division and Death, are outputs of the model. **(B)** Kinetics of the populations of tumor and active dendritic cells, as well as of CALR surface exposure. **(C)** Kinetics of the helper and regulatory CD4$^+$ T cell subpopulations. **(D)** Kinetics of the CD8$^+$ T cell subpopulations.

The model files can be found at: https://github.com/sysbio-curie/ICD, the model description with the meaning of the variables and the parameter values are detailed in **Supplementary Table 2**, and the list of logical rules are provided in **Supplementary Table 3**.

3. RESULTS

To simulate the different steps of ICD, we considered two models: the *Phenomenological* and the *Extended* models. The first model contains ICD markers and a minimal number of cell types. It is easy to handle because of its limited size and serves as a basis for the more detailed version, the *Extended* model.

3.1. Phenomenological Model of ICD

We constructed a simplified model of ICD, focusing on a few key cell types and their interactions (**Figure 2**). This model contains 13 nodes (therefore $2^{13} = 8192$ possible states) and 19 transition rate parameters (see **Supplementary Material**). The nodes of the regulatory network are each associated with a Boolean variable, and transition rates account for the timing of events.

In the regulatory graph shown in **Figure 2A**, the ellipses represent the different cell types considered: Tumor cell, Dying Tumor Cell, Dendritic Cell (DC), Activated Dendritic Cell (ActDC), Migrating Dendritic Cell (MigDC), Lymph Node Dendritic Cell (LNodeDC), T Cell (TCell), and Cytotoxic T Lymphocyte (CTL).

To keep the model as simple as possible, the network contains numerous shortcuts and over-simplifications: for instance, the arc from LNodeDC to TCell in **Figure 2A** represents the fact that a single dendritic cell usually activates a single T Cell.

We introduce different successive status for the dendritic cells (Activated, Migrating, in the Lymph Node), which delay T cell activation. This insures that T cells do not activate immediately after the activation of DC by the ICD-emitted danger signals (CALR, ATP, and HMGB1).

Out of the 19 parameters of this model, 14 must be set manually (the 5 others are updated at the population level and their initial values are set to 0). As the timing of ICD can vary between experimental models, we decided to choose values that have the correct order of magnitude and reproduce the expected timing of ICD (these parameter values are given in **Supplementary Table 1**). More specifically, we set the mean time of "activation" of CALR and ATP at 4 h, and of HMGB1 at 6 h post-ICD inducing chemotherapy (Fucikova et al., 2011). Additionally, we set at "low" the direct cytotoxic efficacy of the treatment, the mean time of DC migration from the tumor to the lymph nodes at 5 days, the mean time of division of tumor cells and CTLs at 10 and 1/2 day(s), respectively, and the mean time of differentiation of T Cells into CTLs at 2 days.

The results of the simulations of the model are shown in **Figures 2B–D**. Following ICD-inducing intervention, the release of CALR, HMGB1, and ATP by dying cancer cells is observable within hours (**Figure 2B**). After 100 h (**Figure 2C**), the immune system is activated with a slow increase of T cells, which peaks at 200 h. The tumor cells are cleared out at about 220 h, which coincides with an increase of the CTL population. To investigate the role of the immune system in the disappearance of the tumor cells, we removed the clonal expansion of the CTLs in the phenomenological model, i.e., the recruitment of immune cells following chemotherapy. In the absence of such increase of the immune effector population, tumor cell clearance becomes less efficient since it relies mostly on the direct cytotoxicity of the treatment (**Figure 2D**).

The phenomenological model successfully reproduces the series of events that are associated with ICD and leading to tumor cell killing following immunogenic chemotherapy. However, the predictive power of this model remains limited because of the lack of molecular details. Hence, we decided to extend this model by further detailing the molecular intermediates, with a focus on the intercellular dialogues.

3.2. Extended Model of ICD

To improve our phenomenological model, we introduced additional nodes representing molecular factors. More precisely, variables such as *MigrDC* and *LNodeDC*, accounting for the activated dendritic cells transiting in the circulation or reaching the tumor draining lymph node, respectively, were replaced by molecules mediating these phenomena.

In this extended version of the model (**Figure 3A**), four populations of cells are considered: tumor cells, dendritic cells, CD4$^+$ and CD8$^+$ T cells. As for the phenomenological model, tumor cells can be converted into dying tumor cells when treated by chemotherapeutic agents, whereas dendritic cells become active after sensing ATP, HMGB1, CALR, and/or ANXA1.

Without treatment, tumor cells proliferate indefinitely. In our model, we implicitly assume that the cells have been treated by an ICD-inducing therapy (node *ChemoT*). Also, we set the initial size of the tumor cell population. These tumor cells produce stress-induced ligands (DAMPs) (Fucikova et al., 2015): CALR, ATP, ANXA1, and HMGB1, whose activation is conditioned by several constraints. For instance, HMGB1 is ready to be released only if ANXA1 is present in the extracellular milieu.

Initially inactive and distant from the tumor bed, DCs become active (denoted by *aDC*) after the stimulation of purinergic receptors upon tumor-derived extracellular ATP binding (mainly the high-affinity metabotropic P2Y2R and the low-affinity ionotropic P2X7R; Rossi et al., 2012), which ignites the migratory status of the DC. This biological information has been translated into a Boolean rule as follows: if a DC and P2X7R are both active at the same time, then DC switches to its active state. Interaction between the activated DC and CALR at the surface of the dying tumor cell triggers its phagocytosis, thereby promoting tumor antigen uptake by the DC (Galluzzi et al., 2017).

Along their way to the lymph node, activated DCs (*aDCs*) capturing antigens upregulate MHC (class-I and II) molecules together with co-stimulatory molecules such as CD80. Once in the secondary, or eventually tertiary, lymphoid tissue, the encounter between such mature DC presenting tumor antigens and a naive undifferenciated T lymphocyte can lead to an activated T cell (Zehn et al., 2012). T cell activation occurs when the cognate TCR and CD28, that are exposed on the lymphocyte, interact with the antigen-loaded MHC and CD80 at the surface of mature DCs, respectively (Galluzzi et al., 2017; Patente et al., 2018). The transition rate associated with the activation of each receptor node is function of the state of the ligand to which it binds. Thus, the activity of the CD8$^+$ T cell receptor (labeled "TCR-I") node is affected by the activity of the MHC-I node. The more MHC-I is produced (i.e., high activation probability), the more likely it is to activate TCR-I.

CXCR3 is a chemotactic receptor on activated T cells, which binds CXCL10 released by tumor cells and by intratumoral activated DCs, following autocrine and paracrine stimulation by type-1 IFNs (**Figure 3**) (Zitvogel et al., 2015; Galluzzi et al., 2017). When CD28 is active, we are considering a single T cell (either CD4 or CD8) whose CXCR3 receptors have already been activated. Further details regarding these interactions are provided in the **Supplementary Figures 1**, **2**.

Activated DCs not only secrete type-1 IFNs but also additional ligands such as the cytokines IL-6 and IL-12, which impact the differentiation of naive T cells as detailed below (**Figure 3**) (Henry et al., 2008; Subbiah et al., 2018).

For the sake of simplicity, we considered a CD4$^+$ T cell as activated (denoted by *cd4_Act* ON) when a lymphocyte expressing both CD28 and a "TCR-II" (i.e., a TCR that recognizes a cognate MHC-II associated with an antigen epitope) interacts with a mature DC (Zehn et al., 2012; Chen and Flies, 2013). Following activation, CD4$^+$ T cells differentiate into effector or regulatory subtypes, depending on the cytokines locally present. Of note, the memory compartment, including the central and effector memory T cell subsets, is not taken into consideration in our current model. Therefore, in the presence of DC-produced IL-12, CD4_Act will differentiate into Th1 CD4$^+$ T cells. By contrast, the sensing of T cell-produced IL-4 induces the Th2 program. Alternatively, the detection of IL-6, TGFb as well as IL1B promotes Th17 differentiation. Also, binding of TGFb engages the undifferenciated lymphocytes into the immunosuppressive Treg lineage (**Figure 3A**).

To account for the fact that Th0 cells do not solely lead to Th1 or Th2 cells, we included Th2, Th17 and Treg subtypes, even though their respective role in the series of immune reactions that follow cancer ICD remains poorly characterized (Galluzzi et al., 2017).

It has been reported that IL-4 is produced by either Th0, Th2 or precursors of $CD8^+$ cytotoxic T lymphocytes (CTLs) (see below for details about $CD8^+$ T cell subtypes) (Zhu, 2015; Farhood et al., 2019). The activation (or release) of this interleukin is thus conditioned by the presence of either cell type.

This means that a naive $CD4^+$ can be driven into proliferation by sensing IL-4, while Th2 is able to sustain its activity.

In parallel, naive $CD8^+$ T lymphocytes can be primed (i.e., co-stimulation of TCR-I and CD28) by mature DCs presenting tumor antigen epitopes onto MHC-I molecules and turned into cytotoxic precursors (*preCTL*). Through the release of both IL-2 and IFNg, Th1 lymphocytes further support the differentiation of preCTL into CTLs (Galluzzi et al., 2017; Farhood et al., 2019). preCTL can be activated by the co-stimulation of MHC-I (represented by the receptor TCR-I in the model) and in the presence of IL-2. They turn into CTL under the influence of IFNg or that of IL-2. While IL-2 is produced by several cell types including mature DCs, Th0, Th1, and Th2, IFNg is only produced by Th0, Th1 and CTLs under the combined effect of several cytokines that enhance its production (Bhat et al., 2017).

CTLs can then release perforin, a cytolytic protein able to form pores in target cells and allow pro-apoptotic proteases to initiate cell death (Halle et al., 2016). Intracellular granules of perforins can be replenished allowing CTLs to kill more than one target tumor cell. If tumor cells can die following a chemotherapeutic treatment, they can also be efficiently eliminated by CTLs excreting perforins. It is modeled here by an amplification factor which ensures a rapid decrease of the remaining tumor cells.

The extended model contains 98 parameters. Among them, 20 are updated at the population level and are initially set to 0. The parameter values are listed into the **Supplementary Table 2**. As for the phenomenological model, not all the parameter values are found in the literature; thus, most of them were set to reasonable values with a correct order of magnitude. Specifically, the mean activation time of proteins have been set to 6 h, except for ATP and CALR, for which this parameter is set to 4 h, according to the literature (Liu et al., 2019; Turubanova et al., 2019; Galluzzi et al., 2020; Humeau et al., 2020), and the mean degradation time of proteins is set to 12 min. We also considered a short activation time of 12 min for the cell types: aDC, Th1, Th2, Th17, Treg, preCTL, as the timing is controlled by proteins. This is not the case for CTL activation, whose mean activation time is in the order of 2.4 h. We consider a very slow growing population of tumor cells (mean time of division is $100,000$ h), and the half-life of the chemotherapeutic agent was estimated at 3 days in the tumor tissue (An and Morris, 2012).

ICD induction was considered to occur 1 h after delivery of the pharmaceutical compound, as it needs $50,000$ h for a tumor to be cleared out by direct cytotoxicity of chemotherapy. We chose a slow growing population of tumor cells and a slow direct cytotoxicity of chemotherapy in order to focus the model on ICD.

During clonal expansion, a lymphocyte needs 20 h to divide. If a tumor cell is completely surrounded by CTLs with perforins, it dies within 1 h. If a CTL with active perforin is completely surrounded by tumor cells, it will loose its perforins in 6 h (but perforin may be reactivated), and will be definitively inactive after 10 days when CTL is estimated to experiment death. These latter parameters are difficult to estimate from experimental data, which justified the sensitivity analysis below.

The simulations of the extended model recapitulated the succession of events leading to ICD. The initial population is composed of 80% of tumor cells, 10% of dendritic cells, and 5% of inactive $CD4^+$ and $CD8^+$ cells. As shown in **Figure 3D**, following the chemotherapeutic treatment, cells start to die, expose CALR on their surface, and release other DAMPs (ATP, ANXA1, HMGB1), ultimately triggering DC maturation. $CD4^+$ and $CD8^+$ cells are then activated and are able to differentiate into T helpers and cytotoxic subtypes. The population of tumor cells undergoes a fast decay starting from 250 h (**Figure 3B**), corresponding to the activation of the adaptive immune response, when Th1 (as well as other subtypes), and most importantly CTLs, are engaged (**Figures 3C,D**). Ultimately, tumor cells are targeted and cleared out by perforins upon CTL degranulation (**Figure 3D**).

3.3. Sensitivity Analysis

We performed a sensitivity analysis to test the global robustness of the extended model. The model contains 98 parameters, but 20 of them have 0 as initial value and are updated at the population level. Therefore, the sensitivity analysis is performed on the 78 remaining parameters. In this respect, we systematically increased and decreased each parameter by 50% separately, which corresponds to 156 model variants (see **Supplementary Table 4**). To assess the effect of the immune system on the killing of the population of tumor cells, we plotted the tumor size at `time = 220 h` and `time = 280 h`, which corresponds to the time frame when the tumor cell population undergoes a swift decrease of its size in the initial "standard" conditions (WT) of the extended model (**Figure 3B**).

The sensitivity analysis shows that the model is quite robust to parameter changes. Indeed, the drop in size of the tumor cell population is slightly anticipated or delayed for a few parameter changes when compared to the WT condition (**Figure 4A**). For instance, the parameters that control the number of DCs are showing the strongest effect (`$InitDC` in WT model, and `More_InitDC` and `Less_InitDC` in the model variants): a lower amount of DCs is delaying time of death whereas a higher amount is accelerating the process. A similar effect is observed for the parameter controlling the rate of T cell clonal expansion (`$clonal_exp_rate` in the WT model, and `More_clonal_exp_rate` and `Less_clonal_exp_rate` in the model variants). Nevertheless, these strongest effects only affect slightly the drop of tumor size (**Figure 4B**). A full list of the effects observed following parameter changes can be found in the **Supplementary Table 4**, where the acronym "More" corresponds to +50%, and "Less" correspond to −50%.

In order to suggest possible points of intervention to further stimulate ICD, we performed the same sensitivity

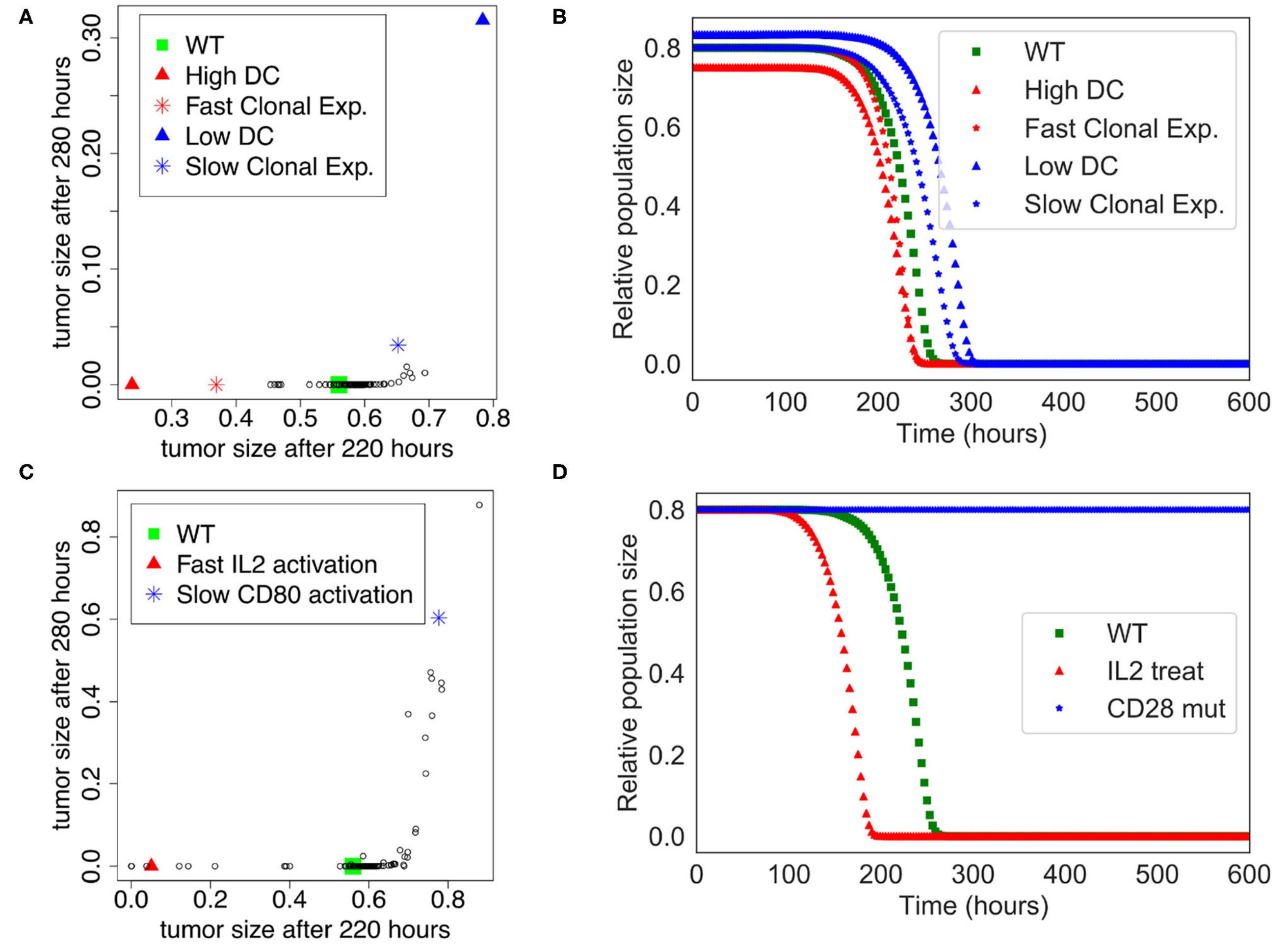


FIGURE 4 | Sensitivity analysis of the Extended Model. **(A)** Tumor size at time= 220 h vs. time= 280 h when the different parameters of the extended model were increased or decreased by 50%. Initial amount of Dendritic cells and rate of clonal expansion show the strongest effect (for WT, $*InitDC*= 0.1, $*clonal_exp_rate*= 0.05; for High DC, $*InitDC*= 0.15 and Low DC, $*InitDC*= 0.067; for Fast Clonal Exp., $*clonal_exp_rate*= 0.075, and for Slow Clonal Exp., $*clonal_exp_rate*= 0.033). **(B)** Kinetics of the size of the tumor cell population according to the parameters changes highlighted in **(A)**. **(C)** Tumor size at time= 220 h vs. time= 280 h when the different parameters of the extended model were multiplied or divided by 5. Faster activation of IL2 and slower activation of CD80 have the strongest effect. **(D)** Kinetics of the size of the tumor cell population following a treatment with IL2 (IL2 treat) or a mutation of CD28 on T cells (CD28 mut), inspired by the parameters changes highlighted in **(C)**.

analysis, but by multiplying and dividing by a factor of 5 each parameter separately (**Figure 4C**), thus mimicking a mutation of the corresponding node. Among the strongest effects, we selected two conditions: the slower activation rate of CD80 and the faster activation rate of IL-2, labeled `LLess_rate_CD80_i` and `MMore_rate_IL2_i`, respectively (cf. **Supplementary Table 5**). These observations led us to simulate a complete knock out of CD28 (CD28 is the target of CD80) and an external treatment by IL-2. The results of these two modifications are shown in **Figure 4D**. In the case of a CD28 knock-out, we observed that 80% of the tumor cell population persists at t=280 h, sign of a failure of the ICD-inducing treatment (knowing that the tumor cell population is initially set to represent 80% of the total number of cells at the beginning of the simulation). Similarly, a treatment that would lead to an increase of IL-2 could kill the tumor cells faster at $t = 200$ h.

This analysis shows that our extended model enables the exploration of perturbations that potentiate the killing of tumor cells by boosting the adaptive immune response which follows an ICD-inducing chemotherapeutic treatment. It is important to note, though, that the approach is not quantitative, and cannot provide regimens of drug treatments, but it can highlight potential mechanisms and molecular targets that could increase tumor clearance.

4. DISCUSSION

ICD of tumor cells is induced by a combination of factors and requires a cooperation between several players of the tumor microenvironment, in particular T cells and DCs. If we still lack a full understanding of the molecular mechanisms governing ICD, as well of the cross-talk between the ICD-induced immune players and the tumor cells, we believe that mathematical modeling could contribute to a more comprehensive understanding of these processes. For this purpose, we integrated information about ICD dispersed in multiple scientific papers into regulatory networks, which explicitly consider how the main cell types communicate.

More specifically, we have constructed two models with two different purposes. The first model aimed at verifying that the series of events that lead to ICD could be reproduced, and at suggesting some parameter values that could mimic the current knowledge related to the timing of these events. This phenomenological model was then used as a basis for a more complex model that included additional key molecules involved in the cascade of events associated with ICD. With the extended model, we were able to further explore the dynamics of the cell type populations subjected to different conditions (number of DCs, speed of some processes, etc.).

To do so, we used a stochastic simulation environment accounting for cell death, cell division and inter-cellular communication to monitor population sizes for different cellular conditions. Interestingly, although certainly still over-simplified, we could recapitulate several essential features of ICD with our model, and even pinpoint the roles of specific components, which might be properly acted upon to boost the immune response.

Such results could be interpreted as possible pharmacological interventions that could improve chemotherapy outcome. The model is an important and necessary tool in such a context because many of these parameters that we can explore with the model are difficult to measure experimentally. To that end, sensitivity analyses have confirmed that the model is suited for strong ICD inducers, like oxaliplatin. To contextualize the model for weak ICD inducers (like mitoxantrone), other mechanisms should be added to the model. The fact that we were only able to switch the timing of complete removal of tumors rather than to reduce the effect of the tumor removal suggests that some essential molecules are still missing in the model.

Future work will include specific *in vitro* and *in vivo* experiments in order to fit parameters to the data and match experimentally-observed timing of the different events leading to tumor clearance. We also plan to extend the model by refining some already described intercellular interactions, e.g., details about the production and effects of IFNg or TGFb on the immune cells, and also by including major signaling pathways inside each cell type to allow more candidates in the search for improving ICD.

A long-term goal is to propose feasible pharmacological interventions that can boost ICD for killing tumor cells, probably by targeting multiple elements throughout the ICD process. The first step of this approach is done with the sensitivity analysis on the refined mathematical model to identify the candidates to target. The second step requires a confirmation of the results with public omics data, a list of potential pharmacological targets, as well as *in vivo* validation, together with further pharmacodynamics/pharmacokinetics studies.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JP, GS, and LC conceived the project. GS, LC, and DT supervised the study. AC constructed the logical models. GK and JP provided expertise in the biological interpretation. AN, EB, and VN provided expertise in the computational biology. GS, LC, AC, and JP wrote the manuscript. AN, VN, EB, GK, and DT edited the manuscript. All co-authors agreed on the manuscript.

FUNDING

JP was supported by the Association Française d'Hépatologie (AFEF); the SIRIC Cancer Research and Personalized Medicine (CARPEM); and the Seerave Foundation. GK was supported by the Ligue contre le Cancer (équipe labellisée); Agence National de la Recherche (ANR)—Projets blancs; ANR under the frame of E-Rare-2, the ERA-Net for Research on Rare Diseases; Association pour la recherche sur le cancer (ARC); Cancéropôle Ile-de-France; Chancellerie des universités de Paris (Legs Poix), Fondation pour la Recherche Médicale (FRM); a donation by Elior; European Research Area Network on Cardiovascular Diseases (ERA-CVD, MINOTAUR); Gustave Roussy Odyssea, the European Union Horizon 2020 Project Oncobiome; Fondation Carrefour; High-end Foreign Expert Program in China (GDW20171100085 and GDW20181100051), Institut National du Cancer (INCa); Inserm (HTE); Institut Universitaire de France; LeDucq Foundation; the LabEx Immuno-Oncology; the RHU Torino Lumière; the Seerave Foundation; the SIRIC Stratified Oncology Cell DNA Repair and Tumor Immune Elimination (SOCRATE); and the SIRIC Cancer Research and Personalized Medicine (CARPEM).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.590479/full#supplementary-material

| GitHub repository at https://github.com/sysbio-curie/ICD/: 1. Jupyter notebook for phenomenological model (ICD_phenomenological.ipynb); 2. Jupyter notebook for the extended model (ICD_extend.ipynb); and 3. model files with .bnd .cfg .upp files.

Supplementary Table 1 | Table of variables and parameters for the phenomenological model: Table_S1_ICD_phenomenological_description.xlsx.

Supplementary Table 2 | Table of variables and parameters for the extended model: Table_S2_ICD_extended_description.xlsx.

Supplementary Table 3 | Table of logical rules: Table_S3_logical_rules.txt.

Supplementary Table 4 | Sensitivity analysis with all parameters increased and decreased by 50%: Table_S4_SensitTum220vs280_0.5.xlsx.

Supplementary Table 5 | Sensitivity analysis with all parameters increased and decreased by 5: Table_S5_SensitTum220vs280_5.xlsx.

Supplementary Figure 1 | Ligand-receptor dynamics related to dendritic cells. DC shows receptors (blue nodes) on its surface (dashed arcs) that can interact with ligands (orange nodes) and can then release ligands (pink nodes).

Supplementary Figure 2 | $CD4^+$ cell dynamics with description of cell differentiated lineages and the role of the cytokines in the activation of the $CD4^+$ T cell dynamics. Black arcs represent activating influences and red arcs inhibiting influences.

REFERENCES

An, G., and Morris, M. E. (2012). A physiologically based pharmacokinetic model of mitoxantrone in mice and scale-up to humans: a semi-mechanistic model incorporating DNA and protein binding. *AAPS J.* 14, 352–364. doi: 10.1208/s12248-012-9344-7

Apetoh, L., Ghiringhelli, F., Tesniere, A., Obeid, M., Ortiz, C., Criollo, A., et al. (2007). Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. *Nat. Med.* 13, 1050–1059. doi: 10.1038/nm1622

Bezu, L., Gomes-da Silva, L. C., Dewitte, H., Breckpot, K., Fucikova, J., Spisek, R., et al. (2015). Corrigendum: "combinatorial strategies for the induction of immunogenic cell death". *Front. Immunol.* 6:275. doi: 10.3389/fimmu.2015.00187

Bhat, P., Leggatt, G., Waterhouse, N., and Frazer, I. H. (2017). Interferon derived from cytotoxic lymphocytes directly enhances their motility and cytotoxicity. *Cell Death Dis.* 8:e2836. doi: 10.1038/cddis.2017.67

Chen, L., and Flies, D. B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* 13, 227–242. doi: 10.1038/nri3405

Dumitriu, I. E., Bianchi, M. E., Bacci, M., Manfredi, A. A., and Rovere-Querini, P. (2007). The secretion of HMGB1 is required for the migration of maturing dendritic cells. *J. Leukoc. Biol.* 81, 84–91. doi: 10.1189/jlb.0306171

Elliott, M. R., Chekeni, F. B., Trampont, P. C., Lazarowski, E. R., Kadl, A., Walk, S. F., et al. (2009). Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature* 461, 282–286. doi: 10.1038/nature08296

Farhood, B., Najafi, M., and Mortezaee, K. (2019). CD8+ cytotoxic T lymphocytes in cancer immunotherapy: a review. *J. Cell. Physiol.* 234, 8509–8521. doi: 10.1002/jcp.27782

Fucikova, J., Kralikova, P., Fialova, A., Brtnicky, T., Rob, L., Bartunkova, J., et al. (2011). Human tumor cells killed by anthracyclines induce a tumor-specific immune response. *Cancer Res.* 71, 4821–4833. doi: 10.1158/0008-5472.CAN-11-0950

Fucikova, J., Moserova, I., Urbanova, L., Bezu, L., Kepp, O., Cremer, I., et al. (2015). Prognostic and predictive value of DAMPs and DAMP-associated processes in cancer. *Front. Immunol.* 6:402. doi: 10.3389/fimmu.2015.00402

Galluzzi, L., Buqu, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2017). Immunogenic cell death in cancer and infectious disease. *Nat. Rev. Immunol.* 17, 97–111. doi: 10.1038/nri.2016.107

Galluzzi, L., Vitale, I., Warren, S., Adjemian, S., Agostinis, P., Martinez, A. B., et al. (2020). Consensus guidelines for the definition, detection and interpretation of immunogenic cell death. *J. Immunother. Cancer* 8:e000337. doi: 10.1136/jitc-2019-000337

Gardai, S. J., McPhillips, K. A., Frasch, S. C., Janssen, W. J., Starefeldt, A., Murphy-Ullrich, J. E., et al. (2005). Cell-surface calreticulin initiates clearance of viable or apoptotic cells through trans-activation of LRP on the phagocyte. *Cell* 123, 321–334. doi: 10.1016/j.cell.2005.08.032

Giovanelli, P., Sandoval, T. A., and Cubillos-Ruiz, J. R. (2019). Dendritic cell metabolism and function in tumors. *Trends Immunol.* 40, 699–718. doi: 10.1016/j.it.2019.06.004

Halle, S., Keyser, K., Stahl, F., Busche, A., Marquardt, A., Zheng, X., et al. (2016). *In vivo* killing capacity of cytotoxic T cells is limited and involves dynamic interactions and T cell cooperativity. *Immunity* 44, 233–245. doi: 10.1016/j.immuni.2016.01.010

Henry, C. J., Ornelles, D. A., Mitchell, L. M., Brzoza-Lewis, K. L., and Hiltbold, E. M. (2008). IL-12 produced by dendritic cells augments CD8+ T cell activation through the production of the chemokines CCL1 and CCL17. *J. Immunol.* 181, 8576–8584. doi: 10.4049/jimmunol.181.12.8576

Humeau, J., Sauvat, A., Cerrato, G., Xie, W., Loos, F., Iannantuoni, F., et al. (2020). Inhibition of transcription by dactinomycin reveals a new characteristic of immunogenic cell stress. *EMBO Mol. Med.* 12:e11622. doi: 10.15252/emmm.201911622

Kepp, O., Menger, L., Vacchelli, E., Adjemian, S., Martins, I., Ma, Y., et al. (2012). Anticancer activity of cardiac glycosides: At the frontier between cell-autonomous and immunological effects. *Oncoimmunology* 1, 1640–1642. doi: 10.4161/onci.21684

Kepp, O., Pol, J., Zitvogel, L., and Kroemer, G. (2018). *Immunogenic Stress and Death of Cancer Cells in Natural and Therapy-Induced Immunosurveillance.* Cham: Springer International Publishing. doi: 10.1007/978-3-319-62431-0_12

Kroemer, G., Galluzzi, L., Kepp, O., and Zitvogel, L. (2013). Immunogenic cell death in cancer therapy. *Annu. Rev. Immunol.* 31, 51–72. doi: 10.1146/annurev-immunol-032712-100008

Kurokawa, T., Oelke, M., and Mackensen, A. (2001). Induction and clonal expansion of tumor-specific cytotoxic T lymphocytes from renal cell carcinoma patients after stimulation with autologous dendritic cells loaded with tumor cells. *Int. J. Cancer* 91, 749–756. doi: 10.1002/1097-0215(200002)9999:9999<::AID-IJC1141>3.0.CO;2-X

Li, M. O., and Rudensky, A. Y. (2016). T cell receptor signalling in the control of regulatory T cell differentiation and function. *Nat. Rev. Immunol.* 16, 220–233. doi: 10.1038/nri.2016.26

Liu, P., Zhao, L., Pol, J., Levesque, S., Petrazzuolo, A., Pfirschke, C., et al. (2019). Crizotinib-induced immunogenic cell death in non-small cell lung cancer. *Nat. Commun.* 10:1486. doi: 10.1038/s41467-019-09838-y

Menger, L., Vacchelli, E., Adjemian, S., Martins, I., Ma, Y., Shen, S., et al. (2012). Cardiac glycosides exert anticancer effects by inducing immunogenic cell death. *Sci. Transl. Med.* 4:143ra99. doi: 10.1126/scitranslmed.3003807

Panaretakis, T., Kepp, O., Brockmeier, U., Tesniere, A., Bjorklund, A. C., Chapman, D. C., et . (2009). Mechanisms of pre-apoptotic calreticulin exposure in immunogenic cell death. *EMBO J.* 28, 578–590. doi: 10.1038/emboj.2009.1

Patente, T. A., Pinho, M. P., Oliveira, A. A., Evangelista, G. C. M., Bergami-Santos, P. C., and Barbuto, J. A. M. (2018). Human dendritic cells: their heterogeneity and clinical application potential in cancer immunotherapy. *Front. Immunol.* 9:3176. doi: 10.3389/fimmu.2018.03176

Riol-Blanco, L., Sanchez-Sanchez, N., Torres, A., Tejedor, A., Narumiya, S., Corbi, A. L., et al. (2005). The chemokine receptor CCR7 activates in dendritic cells two signaling modules that independently regulate chemotaxis and migratory speed. *J. Immunol.* 174, 4070–4080. doi: 10.4049/jimmunol.17 4.7.4070

Rossi, L., Salvestrini, V., Ferrari, D., Di Virgilio, F., and Lemoli, R. M. (2012). The sixth sense: hematopoietic stem cells detect danger through purinergic signaling. *Blood* 120, 2365–2375. doi: 10.1182/blood-2012-04-422378

Saez, P. J., Vargas, P., Shoji, K. F., Harcha, P. A., Lennon-Dumenil, A. M., and Saez, J. C. (2017). ATP promotes the fast migration of dendritic cells through the activity of pannexin 1 channels and P2X7 receptors. *Sci. Signal* 10:eaah7107. doi: 10.1126/scisignal.aah7107

Stoll, G., Caron, B., Viara, E., Dugourd, A., Zinovyev, A., Naldi, A., et al. (2017). MABOSS 2.0: an environment for stochastic Boolean modeling. *Bioinformatics* 33, 2226–2228. doi: 10.1093/bioinformatics/btx123

Stoll, G., Naldi, A., Noël, V., Viara, E., Barillot, E., Kroemer, G., et al. (2020). Upmaboss: a novel framework for dynamic cell population modeling. *bioRxiv [Preprint]*. doi: 10.1101/2020.05.31.126094

Stoll, G., Viara, E., Barillot, E., and Calzone, L. (2012). Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Syst. Biol.* 6:116. doi: 10.1186/1752-0509-6-116

Subbiah, V., Murthy, R., Hong, D. S., Prins, R. M., Hosing, C., Hendricks, K., et al. (2018). Cytokines produced by dendritic cells administered intratumorally correlate with clinical outcome in patients with diverse cancers. *Clin. Cancer Res.* 24, 3845–3856. doi: 10.1158/1078-0432.CCR-17-2707

Turubanova, V. D., Balalaeva, I. V., Mishchenko, T. A., Catanzaro, E., Alzeibak, R., Peskova, N. N., et al. (2019). Immunogenic cell death induced by a new photodynamic therapy based on photosens and photodithazine. *J. Immunother. Cancer* 7:350. doi: 10.1186/s40425-019-0826-3

Vacchelli, E., Ma, Y., Baracco, E. E., Sistigu, A., Enot, D. P., Pietrocola, F., et al. (2015). Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. *Science* 350, 972–978. doi: 10.1126/science.aad0779

Zehn, D., King, C., Bevan, M. J., and Palmer, E. (2012). TCR signaling requirements for activating T cells and for generating memory. *Cell. Mol. Life Sci.* 69, 1565–1575. doi: 10.1007/s00018-012-0965-x

Zhu, J. (2015). T helper 2 (Th2) cell differentiation, type 2 innate lymphoid cell (ILC2) development and regulation of interleukin-4 (IL-4) and IL-13 production. *Cytokine* 75, 14–24. doi: 10.1016/j.cyto.2015.05.010

Zhu, J. (2018). T helper cell differentiation, heterogeneity, and plasticity. *Cold Spring Harb. Perspect. Biol.* 10:a030338. doi: 10.1101/cshperspect.a030338

Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M. J., and Kroemer, G. (2015). Type I interferons in anticancer immunity. *Nat. Rev. Immunol.* 15, 405–414. doi: 10.1038/nri3845

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 17 November 2020
doi: 10.3389/fphys.2020.602844

Calcium Oscillations in Pancreatic α-cells Rely on Noise and ATP-Driven Changes in Membrane Electrical Activity

*Virginia González-Vélez[1], Anthony Piron[2,3] and Geneviève Dupont[3,4]**

[1] Department Basic Sciences, Universidad Autónoma Metropolitana-Azcapotzalco, CDMX, Mèxico, Mexico, [2] ULB Center for Diabetes Research, Faculté de Médecine, Université libre de Bruxelles (ULB), Brussels, Belgium, [3] Interuniversity Institute of Bioinformatics (IB2), Brussels, Belgium, [4] Unit of Theoretical Chronobiology, Faculté des Sciences, Université libre de Bruxelles (ULB), Brussels, Belgium

In pancreatic α-cells, intracellular Ca^{2+} ($[Ca^{2+}]_i$) acts as a trigger for secretion of glucagon, a hormone that plays a key role in blood glucose homeostasis. Intracellular Ca^{2+} dynamics in these cells are governed by the electrical activity of voltage-gated ion channels, among which ATP-sensitive K^+ (K_{ATP}) channels play a crucial role. In the majority of α-cells, the global Ca^{2+} response to lowering external glucose occurs in the form of oscillations that are much slower than electrical activity. These Ca^{2+} oscillations are highly variable as far as inter-spike intervals, shapes and amplitudes are concerned. Such observations suggest that Ca^{2+} dynamics in α-cells are much influenced by noise. Actually, each Ca^{2+} increase corresponds to multiple cycles of opening/closing of voltage gated Ca^{2+} channels that abruptly become silent, before the occurrence of another burst of activity a few tens of seconds later. The mechanism responsible for this intermittent activity is currently unknown. In this work, we used computational modeling to investigate the mechanism of cytosolic Ca^{2+} oscillations in α-cells. Given the limited population of K_{ATP} channels in this cell type, we hypothesized that the stochastic activity of these channels could play a key role in the sporadic character of the action potentials. To test this assumption, we extended a previously proposed model of the α-cells electrical activity (Diderichsen and Göpel, 2006) to take Ca^{2+} dynamics into account. Including molecular noise on the basis of a Langevin type description as well as realistic dynamics of opening and closing of K_{ATP} channels, we found that stochasticity at the level of the activity of this channel is on its own not able to produce Ca^{2+} oscillations with a time scale of a few tens of seconds. However, when taking into account the intimate relation between Ca^{2+} and ATP changes together with the intrinsic noise at the level of the K_{ATP} channels, simulations displayed Ca^{2+} oscillations that are compatible with experimental observations. We analyzed the detailed mechanism and used computational simulations to identify the factors that can affect Ca^{2+} oscillations in α-cells.

Keywords: computational model, ATP-sensitive potassium channels, action potential, Langevin equation, stochastic channel, plasma-membrane Ca^{2+} ATPase

OPEN ACCESS

Edited by:
Jianhua Xing,
University of Pittsburgh, United States

Reviewed by:
Linford Briant,
University of Oxford, United Kingdom
Pei-Chun Chen,
National Cheng Kung University,
Taiwan

****Correspondence:***
Geneviève Dupont
gdupont@ulb.ac.be

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: *07 September 2020*
Accepted: *28 October 2020*
Published: *17 November 2020*

Citation:
González-Vélez V, Piron A and Dupont G (2020) Calcium Oscillations in Pancreatic α-cells Rely on Noise and ATP-Driven Changes in Membrane Electrical Activity.
Front. Physiol. 11:602844.
doi: 10.3389/fphys.2020.602844

INTRODUCTION

Pancreatic islets respond to changes in blood glucose levels so that β-cells secrete insulin when blood glucose is elevated and α-cells secrete glucagon when it is low. Glucagon mobilizes glucose from the liver and when normoglycemia is reestablished, glucagon release from α-cells is suppressed. Extrinsic and intrinsic factors are involved in glucagon secretion (Briant et al., 2016; Wendt and Eliasson, 2020). Individuals with diabetes often show an impaired glucagon secretion that contributes to their hyperglycaemia (D'Alessio, 2011; Gilon, 2020). However, the detailed mechanism by which α-cells regulate glucagon secretion is not fully understood (Yu et al., 2019).

Pancreatic α-cells are electrically excitable and stimulation of glucagon secretion is secondary to repetitive action potential (AP) firing. In a low glucose medium, AP's occur with a frequency of ∼1–3 Hz. Depolarization of the α-cell plasma membrane allows Ca^{2+} to enter through voltage-gated Ca^{2+} channels, which leads to the exocytosis of secretory granules of glucagon. In agreement with this mechanism, electrical stimulation of α-cells leads to an increase in cell membrane capacitance due to granule fusion, a well-known Ca^{2+} dependent process that precedes glucagon release (Barg et al., 2000; Voets, 2000; Göpel et al., 2004; González-Vélez et al., 2012).

Electrical activity in α-cells is thus accompanied by an increase in intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$), which results from the activation of voltage-gated Ca^{2+} channels. Interestingly, this rise in $[Ca^{2+}]_i$ occurs in the form of oscillations with an average frequency of ~ 0.5 min^{-1}, which is much lower than that of the AP's. Parallel measurements of electrical activity and $[Ca^{2+}]_i$ revealed that each oscillation corresponds to a burst of AP's and that the amplitude of the Ca^{2+} increase correlates with the frequency of AP's (MacDonald et al., 2007; Quoix et al., 2009; Le Marchand and Piston, 2010, 2012; Zhang et al., 2013; Kellard et al., 2020). These bursts of electrical activity are separated by quiescent periods during which $[Ca^{2+}]_i$ is close to basal level. Ca^{2+} oscillations are observed in most α-cells, in low or high glucose medium, although they are much reduced in both amplitude and frequency in high glucose. A key characteristic of these oscillations is their irregularity. Their shape, frequency and amplitude are extremely variable, not only from one cell to another but also in the course of time for one individual cell (Kellard et al., 2020). In this study, we investigated the mechanism responsible for the intermittency of electrical activity and thus for the existence of slow, irregular Ca^{2+} oscillations.

Plasma membrane ATP-sensitive K^+ channels (K_{ATP} channels) play a key role in controlling α-cells electrical activity, although the details of this control are still actively debated (Gilon, 2020; Zhang et al., 2020). When the amplitude of this current is relatively limited, voltage-gated Na^+ and Ca^{2+} channels can indeed initiate an AP. The number of K_{ATP} channels simultaneously active is however surprisingly low, of the order of 10 (Rorsman et al., 2014). With such a low number of channels, it is expected that fluctuations of molecular origin would play a key role in the dynamical evolution of the K_{ATP} current and thus of the whole voltage and Ca^{2+} dynamics (Gonze et al., 2018). The low K_{ATP} channel activity in α-cells also results in a very high input resistance, meaning that small currents as those associated with openings of a few ion channels may have a drastic effect on membrane voltage and electrical activity (Rorsman et al., 2014). In agreement with this reasoning, noise-induced APs have been observed in α-cells and theoretically simulated (Diderichsen and Göpel, 2006). Here, we pushed this observation forward and investigated the possibility that fluctuations related to the small K_{ATP} current might be responsible for the intermittent character of electrical activity and hence for the noisy Ca^{2+} oscillations. This hypothesis, which is ideally investigated by mathematical modeling, holds with the observations that α-cells activity is highly variable even at a given external glucose concentration and that there is no "typical" α-cell signature (Kellard et al., 2020). Importantly, we here focus on the mechanism responsible for the existence of slow, irregular Ca^{2+} oscillations and not on the actively debated mechanism of regulation of glucagon secretion. Regulation of glucagon secretion indeed involves both intrinsic mechanisms and paracrine signals. Here, we only take into account intrinsic processes. Since intermittent electrical activity and slow $[Ca^{2+}]_i$ oscillations have also been observed in *isolated* α-cells (Salehi et al., 2006; Quoix et al., 2009; Tuduri et al., 2009; Le Marchand and Piston, 2010), we indeed reasoned that paracrine signaling is not essential for their existence although it affects their characteristics (Briant et al., 2018b; Kellard et al., 2020).

Mathematical models have been developed to study α-cells electrical activity and glucagon secretion. Diderichsen and Göpel (2006) used experimental data from patch clamp experiments on pancreatic α-cells located on the surface of intact mouse islets to develop an accurate model of plasma membrane electrical activity. This model was extended to include first Ca^{2+} dynamics and secretion (Watts and Sherman, 2014), and later glucagon-like peptide 1 (GLP-1) and adrenaline effects (Montefusco and Pedersen, 2015) as well as the α-cell heterogeneity by introducing realistic cell-to-cell variations in the values of the parameters (Montefusco et al., 2020). A functional identification of the islet cell types based on their electrophysiological characteristics allowed to improve the agreement between experiments and simulations of these models (Briant et al., 2017). Diderichsen and Göpel's model was also re-used by Grubelnik et al. (2020) to study the link between the deformities in mitochondrial ultrastructure observed in α-cells of type 2 diabetes mellitus mice and glucagon secretion. On the other hand, Fridlyand and Philipson (2012) adapted a model initially developed to describe β-cells dynamics to propose a detailed description of α-cells electrical activity, Ca^{2+} changes, metabolism as well as paracrine and endocrine regulations. The effect of paracrine signaling (Watts et al., 2016; Briant et al., 2018b) on α-cells electrical activity was also investigated in models of pancreatic islets including β- and δ-cells. None of these studies have addressed the possible impact of the low number of K_{ATP} channels, nor the question of the mechanistic origin of cytosolic Ca^{2+} oscillations in α-cells.

The present study is based on the original Diderichsen and Göpel's model of α-cells electrical activity. This core model is sequentially extended to take into account Ca^{2+} dynamics and random fluctuations of the K_{ATP} current via the Langevin formalism. We found that stochasticity in this current can indeed induce intermittent electrical activity and Ca^{2+} oscillations, but only for unrealistically small values of the opening and closing rates of this K^+ channel. This theoretical prediction motivated us to further extend the model to take into account the variations of ATP concentrations that result from the activity of the plasma membrane Ca^{2+} ATPases. The resulting changes in ATP, which have been observed experimentally (Li et al., 2015), indeed slow down the dynamics of the K_{ATP} current and allow for intermittent electrical activity and Ca^{2+} oscillations resembling those observed experimentally. Finally, we used the model to investigate the sensitivity of calcium and electrical activities to key factors such as the number of K_{ATP} channels or the rate of Ca^{2+} pumping.

MODEL DESCRIPTION

We base our model (**Figure 1**) on the mathematical description of the electrical activity of pancreatic α-cells proposed by Diderichsen and Göpel (2006), which was carefully calibrated on experimental data (see also Briant et al., 2017). The original model incorporates an ATP-sensitive K^+ current (I_{KATP}) that couples the level of external glucose to the electrical properties of the α-cell plasma membrane, through the sensitivity of the intracellular ATP/ADP ratio to external glucose concentration. It also describes a voltage-gated Na^+ current (I_{Na}), a delayed rectifying (I_{KDR}) and a A-type K^+ current (I_{KA}), an unspecific leak current (I_{leak}) and a L-type (I_{CaL}) and a T-type Ca^{2+} current (I_{CaT}). We also consider an additional type of Ca^{2+} current, because N-type (Quesada et al., 2008; González-Vélez et al., 2010) or P/Q type (Rorsman et al., 2014) Ca^{2+} currents have been shown to play a key role in glucagon secretion in rodents. Here, we arbitrarily chose to incorporate a N-type Ca^{2+} current (I_{CaN}). Ca^{2+} concentrations are described differently just below the plasma membrane and in the cytoplasm. Ca^{2+} entry via the L-, T- and N-type Ca^{2+} channels increases Ca^{2+} concentration in a hypothetical sub-plasma membrane compartment. From this compartment, Ca^{2+} diffuses into the cytoplasm or leaves the cell via the plasma membrane Ca^{2+} ATPase (PMCA). Ca^{2+} efflux from the cell is thus accompanied by the hydrolysis of ATP into ADP. The resulting decrease in ATP concentration provokes an increase in the K_{ATP} conductance, thereby providing a feedback from Ca^{2+} changes on the electrical properties of the membrane. The main features of the model are described here below. Additional information about the equations and the values of the parameters can be found in the **Supplementary Data** and in the original study of Diderichsen and Göpel (2006).

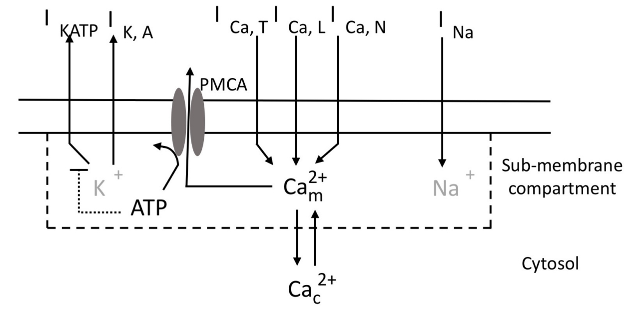


FIGURE 1 | Schematic representation of ionic currents, Ca^{2+} fluxes and ATP consumption in pancreatic α-cells. Transmembrane currents are: I_{KATP}, the ATP-sensitive K^+ current; I_{KA}, the high-voltage activated K^+ current; I_{CaL}, the high-voltage activated L-type Ca^{2+} current; I_{CaT}, the low-voltage activated T-type Ca^{2+} current; I_{CaN}, the high-voltage activated N-type Ca^{2+} current; I_{Na}, the voltage-gated Na^+ current. The model describes the evolution of Ca^{2+} concentration in a fictitious sub-membrane compartment (Ca_m^{2+}) and in the cytosol (Ca_c^{2+}). Exchanges between these 2 compartments occur by diffusion. Ca^{2+} is transported from the sub-membrane compartment into the extracellular medium by the ATP-consuming PMCA. ATP increases in the cytosol at a rate that depends on the concentration of extracellular glucose. ATP inhibits the ATP-sensitive K^+ current (I_{KATP}). K^+ and Na^+ concentrations (in gray) are not explicitly considered in the model.

Model of Plasma Membrane Electrical Activity

Electrical activity is described by the following differential equation:

$$\frac{dV}{dt} = -(I_{CaT} + I_{CaL} + I_{CaN} + I_{Na} + I_{KA} + I_{KDR} + I_{KATP} + I_{leak})/C_m \tag{1}$$

where V is the membrane voltage and C_m is the membrane capacitance set to 5 pF (Diderichsen and Göpel, 2006). For all currents, except for I_{KATP} (see section " ATP-Sensitive K^+ Current and ATP Evolution" below), we use the model proposed by Diderichsen and Göpel (2006) that is based on a Hodgkin-Huxley type description of ion channels and where the activation and inactivation functions have been fitted to experimental data. The high-voltage gated N-type Ca^{2+} current is described as in other studies (Csercsik et al., 2010; González-Vélez et al., 2010).

Ca^{2+} Dynamics

The evolution of sub-membrane Ca^{2+} concentration, noted Ca_m, is modeled as

$$\frac{dCa_m}{dt} = f_r \cdot \frac{-I_{CaT} - I_{CaL} - I_{CaN}}{2 \cdot F \cdot Vol/20} \cdot 10^6 + V_b - V_p \frac{Ca_m^2}{Ca_m^2 + K_m^2} \cdot f_1(ATP) - \gamma(Ca_m - Ca_c) \tag{2}$$

where f_r is the fraction of unbuffered Ca^{2+}, F the Faraday constant and Vol, the volume of an α-cell. It is assumed that the sub-membrane shell represents 1/20 of the total volume of the cytoplasm. The factor 10^6 allows to get concentrations expressed in μM. V_b stands for a basal rate of Ca^{2+} entry, which ensures a ∼0.1 μM Ca^{2+} concentration when voltage-gated Ca^{2+} channels are closed. The third term represents Ca^{2+} pumping out of the cell through PMCA, which can be modeled by a Hill function with $n_H = 2$ (Dupont et al., 2016). This active transport is accompanied by ATP hydrolysis. Thus,

$$f_1(ATP) = \frac{ATP}{ATP + K_e} \tag{3}$$

with K_e being the Michaelis-Menten constant of ATP hydrolysis by the PMCA.

Finally, the last term of Eq. 2 represents Ca^{2+} diffusion into the bulk of the cytoplasm. Given a Ca^{2+} diffusion coefficient of 13 $\mu m^2/s$ (Allbritton et al., 1992) and a diameter of 8 μm for an α-cell (Zimny and Blackard, 1975) in which the nucleus occupies ~70% of the cytoplasm, γ can be estimated to ~0.01 ms^{-1} (see **Supplementary Material**).

The evolution of intracellular ATP follows

$$\frac{dATP}{dt} = V_{gly}\frac{Glu}{Glu + K_{glu}} - \frac{V_p}{2}\frac{Ca_m^2}{Ca_m^2 + K_m^2} \cdot f_1(ATP) - k \cdot ATP \tag{4}$$

The first term of Eq. 4 is a phenomenological description of glycolysis in the form of Michaelis-Menten rate of ATP production from glucose. The second term represents ATP consumption by plasma membrane PMCA and is thus the same as in Eq. 2 except for the factor 2 that takes into account that one mole of ATP is hydrolyzed for two moles of Ca^{2+} transported. Rate constant k describes the consumption of ATP by other intracellular processes.

In agreement with Eq. 1, the evolution of cytoplasmic Ca^{2+} concentration is given by

$$\frac{dCa_c}{dt} = \frac{\gamma}{19}(Ca_m - Ca_c) \tag{5}$$

ATP-Sensitive K$^+$ Current and ATP Evolution

In the original model (Diderichsen and Göpel, 2006), the ATP-sensitive K^+ current is modeled as an ohmic ionic current with a constant conductance, i.e.,

$$I_{KATP} = g_{KATP}(V - V_K) \tag{6}$$

where V_K is the reversal potential for currents carried by potassium. The channel conductance, g_{KATP}, was considered as a constant parameter and an increase in the extracellular glucose concentration was simulated by a decrease in the value of this parameter.

The main goal of the present study is to assess the effect of noise on the K_{ATP} current that arises in α-cells due to the small number of such channels. g_{KATP} is thus not considered as a constant parameter anymore. Instead, we stochastically simulate opening and closing of these channels on the basis of the Langevin formalism. These channels can flicker between an open and a closed state, with opening and closing rate constants denoted α and β, respectively. Considering a noise term, the deterministic evolution equation for the fraction of open channels (s) becomes:

$$\frac{ds}{dt} = \alpha(1 - s) \cdot f_2(ATP) - \beta s + \sqrt{\sigma}\xi(t) \tag{7}$$

where $\xi(t)$ is a random function of time. The last term is a noise function with 0 mean and

$$\sigma = \frac{\alpha(1 - s) \cdot f_2(ATP) + \beta s}{N_{KATP}} \tag{8}$$

with N_{KATP} the number of potentially openable K_{ATP} channels (Fall et al., 2005).

$f_2(ATP)$ takes into account that K_{ATP} channels are reversibly inhibited by ATP (Enkvetchakul et al., 2001). We chose the simple expression:

$$f_2(ATP) = \frac{K_{inh}}{ATP + K_{inh}} \tag{9}$$

where K_{inh} is a constant representing the concentration of ATP leading to half-maximal inhibition.

When taking into account the stochasticity of the K_{ATP} channels, the g_{KATP} conductance appearing in Eq. 6 is now computed as

$$g_{KATP} = g_{KATP}^{s} \cdot s \cdot N_{KATP} \tag{10}$$

with g_{KATP}^{s} standing for the unitary conductance of a single ATP-sensitive K^+ channel.

Together with the equations of the original model (Diderichsen and Göpel, 2006) listed in the **Supplementary Data**, Eqs 1–10 constitute our computational model that has been integrated in Matlab, using an Euler integration scheme with Δt = 0.012 ms. Bifurcation diagrams have been established using the AUTO package of XPPAUT (Ermentrout, 2002).

Outline of the Modeling Approach

As explained in detail in the section "Results," we consider models of increasing complexity to investigate the possible impact of the fluctuations in the K_{ATP} conductance on the existence of Ca^{2+} oscillations in α-cells.

Model 1 is the model proposed by Diderichsen and Göpel (2006), including N-type Ca^{2+} channels and two additional variables: subplasmalemmal Ca^{2+} (Ca_m, Eq. 2 with f_1 = 1) and cytosolic Ca^{2+} (Ca_c, Eq. 5). It is thus defined by Eqs 1, 2, and 5.

Model 2 moreover includes stochasticity in the K_{ATP} conductance through Eqs 7, 8 and 10 that are considered in addition to the equations defining Model 1. f_1 and f_2 are taken equal to 1.

Model 3 allows to investigate the effect of the Ca^{2+}-induced variations in ATP concentration. It is similar to Model 2, except that it includes Eq. 4 to describe the evolution of [ATP]. Changes in the concentration of this nucleotide impact on the evolutions of subplasmalemmal Ca^{2+} via f_1 that is now given by Eq. 3 and of the K_{ATP} channel via f_2 that is now given by Eq. 9.

RESULTS

Calcium Changes Induced by Electrical Activity

When α-cells are electrically active, $[Ca^{2+}]_i$ first rises just beneath the plasma membrane since Ca^{2+} is entering through voltage-gated channels located in this membrane. Sub-plasmalemmal Ca^{2+} is immediately buffered and diffuses in the cytosol where Ca^{2+} concentration is low at rest (~100 nM). This attenuates Ca^{2+} increases below the membrane and transmits signaling to the rest of the cell. Additionally, sub-membrane Ca^{2+} is pumped out of the cell by Ca^{2+} ATPases. *Model 1* takes these fluxes into account and combines a description of α-cell electrical activity with equations for the evolution of subplasmalemmal (Ca_m, Eq. 2) and cytosolic Ca^{2+} (Ca_c, Eq. 5). Simulations of *Model 1* show that each AP generates one spike of Ca^{2+} in the subplasmalemmal space (**Figures 2A,B**). Considering that this fictitious compartment occupies 1/20 of the α-cell volume, corresponding to a shell thickness of ∼100 nm, Ca^{2+} increases up to a level between 0.5 and 1 μM depending on the value of g_{KATP}. A supra-threshold increase in g_{KATP} induces repetitive changes in Ca_m, whose frequency and amplitude are fixed by the value of this conductance. Because of the relative slowness of diffusion, cytosolic Ca^{2+} can only increase during repetitive AP's (**Figure 2C**). When AP's are sustained long enough, cytosolic Ca^{2+} remains nearly constant at a steady level (not shown). Ca^{2+} concentration in the cytosol does not exceed 0.4 μM. In this model, repetitive electrical activity always leads to a sustained increase in cytosolic Ca^{2+}, which does not correspond to the slow Ca^{2+} oscillations observed experimentally in most α-cells.

The main characteristics of Ca^{2+} dynamics for different values of the K_{ATP} conductance are visible in the bifurcation diagram shown in **Figure 2D**. The shape of this diagram is most easily understood by looking at the companion bifurcation diagram showing how electrical activity depends on g_{KATP} (**Figure 2E**). Since Ca^{2+} concentration does not feedback on the cell electrical activity in *Model 1*, Ca^{2+} changes can be seen as a simple output of the latter activity. Starting from large values of the K_{ATP} conductance at which cells are hyperpolarized (right part of the diagram), a decrease in this conductance leads to depolarization and electrical firing, as observed in experiments upon the addition of the K_{ATP} channel blocker tolbutamide. A further decline in the K_{ATP} conductance abolishes electrical activity and the cell remains in a constantly depolarized state. This decline in activity allows the model to reproduce the observation that α-cells are electrically inactive in the presence of large concentrations of external glucose without considering paracrine signaling.

The bifurcation analysis allows to uncover the existence of a range of values of g_{KATP} for which multiple stable solutions can co-exist. This implies that different steady state solutions can be reached depending on the initial conditions. From 0.2265 to 0.2919 nS (blue region in **Figures 2D,E**), the system can potentially be in three states: a stable hyperpolarized state, an intermediate stable state (∼ −35 mV) or an electrically active state with large amplitude AP's. However, the intermediate state, although stable from a physical point of view, is most probably of little physiological significance, as its basin of attraction is very limited. This implies that this state will rarely be reached, and that if reached, fluctuations of internal or external origin would push the system from this intermediate state to the hyperpolarized one or to the limit cycle corresponding to repetitive AP's. For larger values of g_{KATP} (from 0.2919 to 0.3715 nS, green region in **Figures 2D,E**), the labile, slightly polarized state coexist with the stable hyperpolarized one.

Figure 2C shows that this simple model predicts that when cells are in a stationary regime of repetitive AP's, cytosolic Ca^{2+} tends to reach a steady state, although very small changes of concentration occur at each AP (∼20 nM). Such rapid and small changes are not expected to be visible experimentally.

In conclusion, slow Ca^{2+} oscillations resulting from intermittent electrical activity as observed in α-cells cannot result from repetitive AP's controlled by a constant K_{ATP} conductance.

Randomness in the K_{ATP} Conductance

During electrical activity, the number of simultaneously active K_{ATP} channels is very small, of the order of 10 (Rorsman et al., 2014). Thus, the fluctuations related to internal noise cannot be neglected and the deterministic description of their contribution to membrane current (Eq. 6) must be replaced by a stochastic description. We thus considered the transitions of the channel between an open and a closed state, with a noise on this process as described above (*Model 2* in section "Model of Plasma Membrane Electrical Activity"). Rate constants of channel opening and closing (α and β in Eq. 7) estimated from dwell time distributions are of the order of tenths of ms (Enkvetchakul et al., 2001). Assuming a unitary conductance (g^s_{KATP}) of 41 pS (Bokvist et al., 1999; Khan et al., 2001), the total number of K_{ATP} channels in the simulated cell needs to be around 60 to get an average of 5–10 simultaneously open channels during electrical activity, which corresponds to a global K_{ATP} conductance in the range 0.2–0.25 nS. This number fits in the large range of values estimated experimentally (Huang et al., 2011).

Simulations including noise at the level of K_{ATP} current display AP's that are variable in amplitude and frequency. When the average value of g_{KATP} is in the oscillatory range determined by the bifurcation diagram (**Figures 2D,E**), the simulated cell is electrically active (average cellular g_{KATP} = 0.14 nS in **Figures 3E–H**), while the cell is in a fluctuating hyperpolarized state when the average value of g_{KATP} corresponds to a stable steady state of the bifurcation diagram (average cellular g_{KATP} = 0.47 nS in **Figures 3A–D**). Model simulations never result in Ca^{2+} oscillations whatever the values taken for the unitary conductance of the K_{ATP} channel (g^s_{KATP}) and the total number of such channels (N_{KATP}) considered in the simulations. However, we observed that Ca^{2+} oscillations resembling experimental observations could be obtained in the simulations when ascribing to the rate constants of opening and closing of the K_{ATP} channels (parameters α and β) values at least 500 smaller than those reported from experiments (**Figures 3I–L**). In this case, an intermittent electrical activity generates bursts of AP's, with each AP leading to a sharp Ca^{2+} increase in sub-membrane Ca^{2+}. Sub-membrane Ca^{2+} diffuses in the cytoplasm, and because diffusion is slow, one burst of electrical activity involving multiple AP's leads to one Ca^{2+} peak in the cytoplasm. Simulations shown in panels I–K display strong resemblance with experimental observations (MacDonald et al., 2007; Quoix et al., 2009; Le Marchand and Piston, 2010, 2012; Zhang et al., 2013; Kellard et al., 2020). The correlation between the cytosolic Ca^{2+} spike and the period of electrical activity has been shown in isolated mouse α-cells (see for example Quoix et al., 2009) and in α-cells from intact islets (Kellard et al., 2020).

From a mechanistic point of view, when the rate constants of opening and closing of the K_{ATP} channels are small, K_{ATP} current variations last long enough to induce a change in the electrical properties of the membrane: when the cell is hyperpolarized,

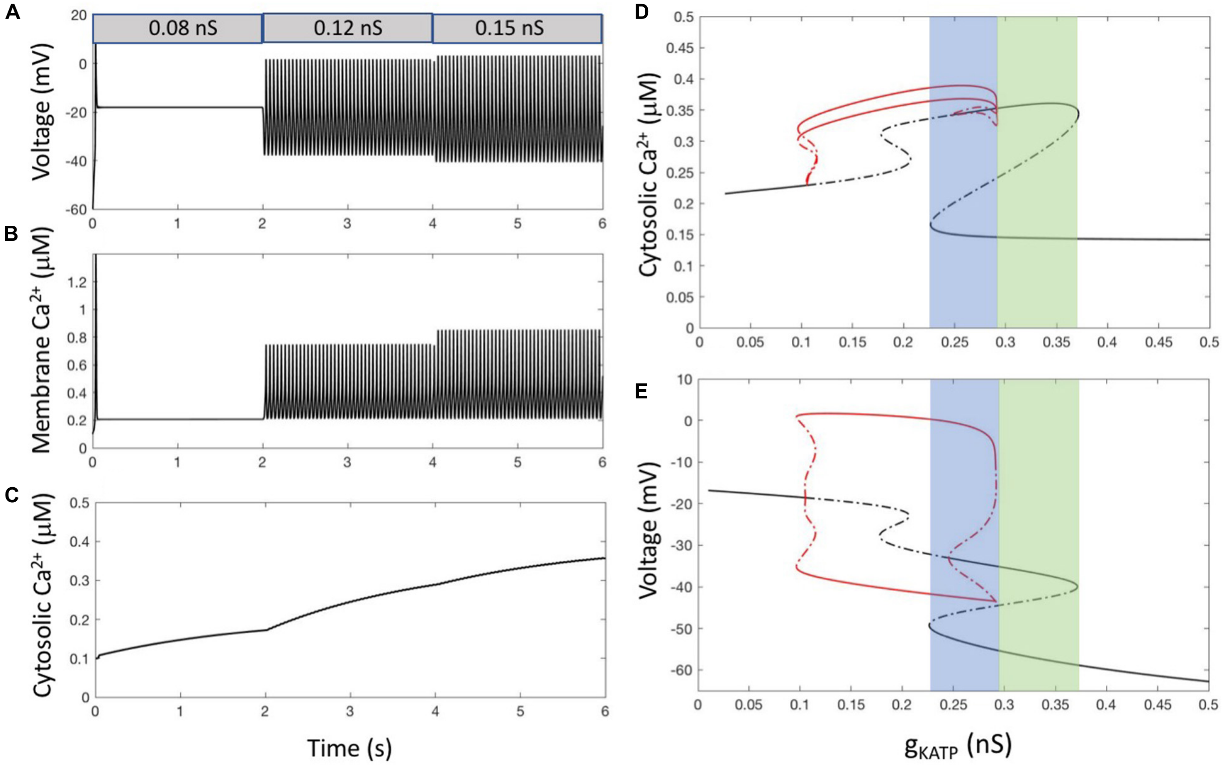


FIGURE 2 | α-cell Ca^{2+} dynamics resulting from electrical activity. **(A)** Membrane voltage for 3 values of g_{KATP}, simulating a decrease in external glucose. From $t = 0$ to 2 s, $g_{KATP} = 0.08$ nS and $g_{leak} = 0.2$ nS. From $t = 2$ to 4 s, $g_{KATP} = 0.12$ nS and $g_{leak} = 0.2$ nS. For $t > 4$ s, $g_{KATP} = 0.15$ nS and $g_{leak} = 0.13$ nS. The decrease in g_{leak} that accompanies the increase in g_{KATP} allows the model to reproduce the experimentally observed lower maximum voltage reached during repetitive AP's at high glucose (Montefusco and Pedersen, 2015). **(B)** Corresponding changes in $[Ca^{2+}]$ in the sub-plasmalemmal compartment (Ca_m). **(C)** Corresponding changes in $[Ca^{2+}]$ in the cytosol (Ca_c). Because diffusion is much slower than electrical activity, cytosolic concentration does not follow the Ca^{2+} spikes occurring just below the membrane. **(D,E)** Bifurcation diagrams showing the evolution of Ca_c **(D)** and voltage **(E)** when changing g_{KATP}. In both panels, black lines indicate steady states, while red lines show the minimum and maximum values reached during limit cycle oscillations. Plain lines indicate stable solutions and dashed lines instable ones. Bifurcation diagrams have been established using the AUTO package of xppaut (Ermentrout, 2002). For all panels, equations are those corresponding to *Model 1* (see section "Outline of the Modeling Approach") with the default values of parameters listed in **Supplementary Tables 1, 2** of the **Supplementary Material**, except for the indicated values of g_{KATP} and g_{leak}. In panels **(D,E)**, $g_{leak} = 0.2$ nS.

a random decrease in g_{KATP} of sufficient duration leads to a cell membrane depolarization that triggers electrical activity and Ca^{2+} entry. Similarly, when the cell membrane is electrically active, a random increase in g_{KATP} of sufficient duration leads to a decrease in V that brings the cell in a resting state. With the experimentally reported values of channel opening and closing rates, the random changes in g_{KATP} are too fast to induce AP's. When simulating a stepwise decrease in g_{KATP} from 0.47 to 0.2 nS in the absence of noise, the change in g_{KATP} must last at least 110 ms to induce electrical activity (not shown). This agrees with the numerical observation that if the fluctuation-driven changes of g_{KATP} have a characteristic time of a few ms as measured experimentally (Enkvetchakul et al., 2001), they will not induce AP's (**Figures 2A–C**).

Thus, simulations predict that Ca^{2+} oscillations may in principle arise from noise-induced changes in the cell K_{ATP} current. However, the time scale of these changes must be much slower than that of the intrinsic dynamics of the voltage-gated ionic channels generating the AP's. This does not correspond to the reported rates of opening and closing of the K_{ATP} channels and raises the possibility that some physiological process drives slow changes in the opening of the K_{ATP} channels.

Ca^{2+}-Driven [ATP] Variations Are Responsible for Slow Changes in Electrical Activity

K_{ATP} channels are regulated by variations of the intracellular ATP concentration (Bokvist et al., 1999). On the other hand, α-cells show oscillations in Ca^{2+} and ATP submembrane concentrations when observed in constant hypoglycemic conditions (Li et al., 2015). These oscillations are in opposite phase, most probably because Ca^{2+} transport out of the cell is an ATP-consuming process. Here, we investigate the hypothesis that even at constant external glucose, ATP/ADP changes may trigger changes in membrane electrical activity and Ca^{2+} entry that would thus be responsible for the observed, slow cytosolic Ca^{2+} oscillations. To this end, we consider [ATP] as a variable in *Model 3*, as well as its relationship with Ca^{2+} dynamics and its inhibitory effect on K_{ATP} channels conductance. For sake of simplicity,

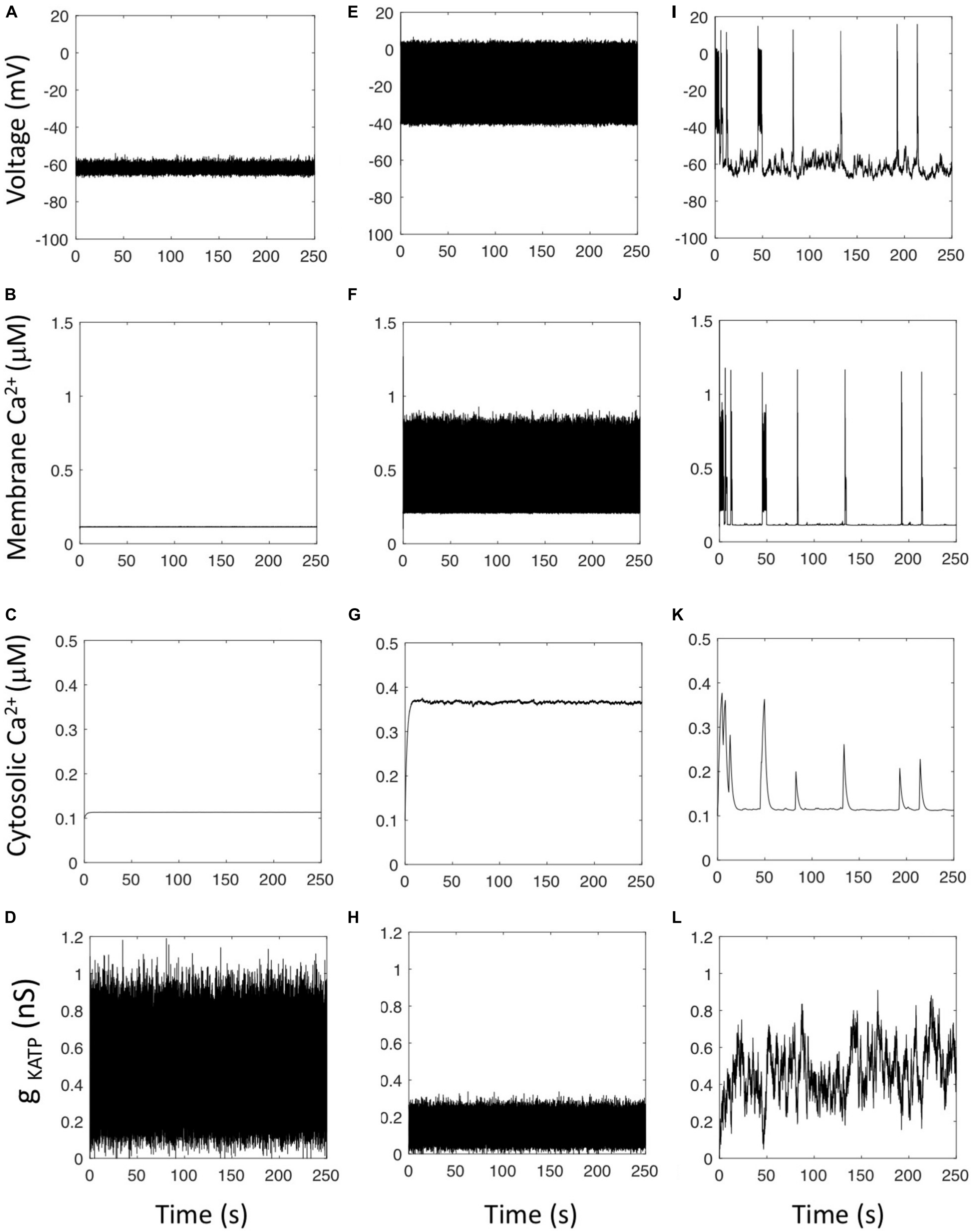


FIGURE 3 | Effect of stochasticity in the K_{ATP} current. Panels show the evolution of the main variables of the model when the conductance of the K_{ATP} channel is described by a Langevin equation (Eqs 7 and 8 with $f_2 = 1$) to take into account the randomness due to the small number of such channels in α-cells. **(A–D)** Rate constant of opening of the K_{ATP} channel, $\alpha = 0.1\ ms^{-1}$, rate constant of closing of the K_{ATP} channel, $\beta = 0.4\ ms^{-1}$; unitary conductance of the K_{ATP} channel$g^{s}_{KATP} = 0.041$ nS. The cell is in a quiet, hyperpolarized state despite the large variations of the total cell g_{KATP} conductance. **(E–H)** When the unitary conductance of the K_{ATP} channel equals 0.012 nS, the cell is in a depolarized state resulting in a constantly high Ca^{2+} concentration. **(I–L)** When considering slow rates of opening and closing of the K_{ATP} channel, $\alpha = 1\ 10^{-4}\ ms^{-1}$ and $\beta = 4\ 10^{-4}\ ms^{-1}$, simulations predict the occurrence of intermittent electrical activity leading to irregular Ca^{2+} spikes in the cytosol. Equations are those corresponding to Model 3 (see section "Outline of the Modeling Approach") with the default values of parameters listed in the **Supplementary Material**, except for the indicated values of g^{s}_{KATP} and $g_{leak} = 0.2$ nS.

we consider that the conductance of these channels depends on [ATP] via an inhibitory function of the Michaelis-Menten type (Eq. 9). Keeping realistic values for the rates of opening and closing of the K_{ATP} channels (parameters α and β), simulations show highly variable Ca^{2+} oscillations that result from intermittent electrical activity (**Figure 4**). Each burst of electrical activity triggers a massive entry of Ca^{2+} in the subplasmalemmal compartment (Ca_m, **Figure 4B**), which then invades the cytoplasm (**Figure 4C**). Because PMCA are fully active, ATP is consumed, and the evolution of its concentration is the mirror image of that of cytosolic Ca^{2+} (**Figure 4D**). This reduced level of ATP favors a large K_{ATP} current, which in turn reduces electrical activity and Ca^{2+} entry, thus allowing ATP to rise again.

Although fluctuations of the K_{ATP} conductance are very rapid (**Figure 4E**), g_{KATP} is in average larger when [ATP] is low (periods of activity) than when [ATP] is high (quiescent periods) because of the inhibition of the K_{ATP} current by ATP. This is visible when computing the moving average of g_{KATP} (**Figure 4F**). The trend in g_{KATP} evolution indeed correlates with that of ATP. Such a trend in the moving average of g_{KATP} does not appear when ATP is considered as a constant (*Model 2*, not shown). As a consequence of these trends, large, unlikely perturbations are necessary to switch from an inactive to an active period and *vice-versa*. Thus, the mutual interaction between ATP dynamics and electrical activity that occurs through the Ca^{2+} changes creates trends in the changes in electrical activity allowing for intermittent activity, despite the rapid fluctuations in g_{KATP} .

The cytosolic Ca^{2+} oscillations shown in **Figure 4C** have widely different durations, from ∼15 s to ∼2 min. Because of stochasticity, these values are different from one simulation to the other. Given that glucagon secretion is triggered by Ca^{2+} increases above a certain threshold (González-Vélez et al., 2012), it is interesting to investigate what controls the ratio of active *versus* inactive periods. As randomness is required to initiate changes from an active to a quiescent period and *vice-versa*, it can be expected that the number of K_{ATP} channels will play a key role in controlling the number of transitions and thus the activity ratio. To investigate the effect of this factor, we performed simulations with different values of N_{KATP} and computed the fraction of time during which cells exhibit electrical activity (**Figure 5C**) as well as the average number of Ca^{2+} spikes during a 500 s simulation (**Figure 5E**). In the simulations, changes in the

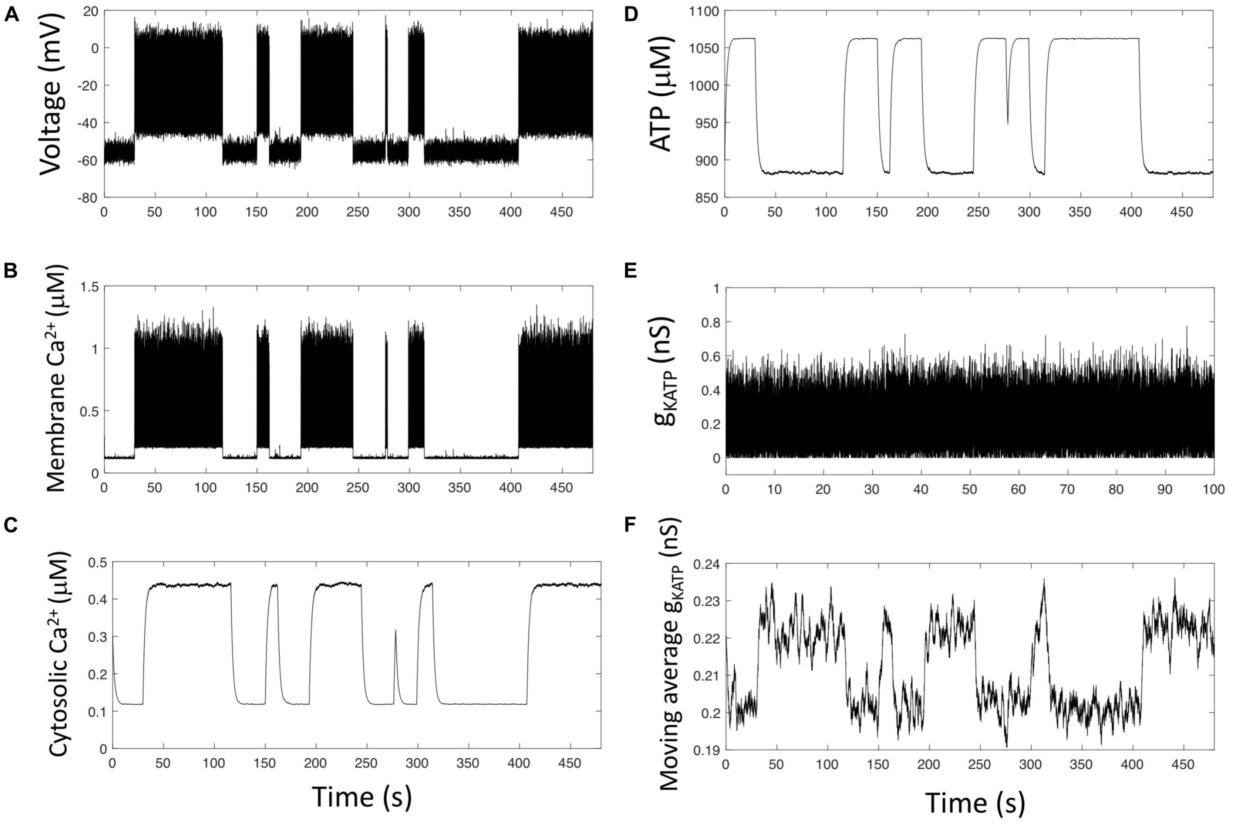


FIGURE 4 | Simulations predict that changes in membrane ATP and stochasticity in the K_{ATP} current generate slow Ca^{2+} oscillations. **(A–E)** Numerical simulations of *Model 3* (see section "Outline of the Modeling Approach") that describes the opening and closing of the K_{ATP} channel by a Langevin stochastic equation and takes into account the consumption of ATP by PMCA as well as the inhibition of K_{ATP} channels opening by ATP. **(F)** The moving average of g_{KATP} allows to visualize that, although highly random, g_{KATP} tends to increase when ATP concentration is low and to decrease when ATP concentration is large. There are thus long-lasting trends in g_{KATP} changes that can induce robust changes in α-cells electrical activity. Values of g_{KATP} are averaged on a 200 s period.

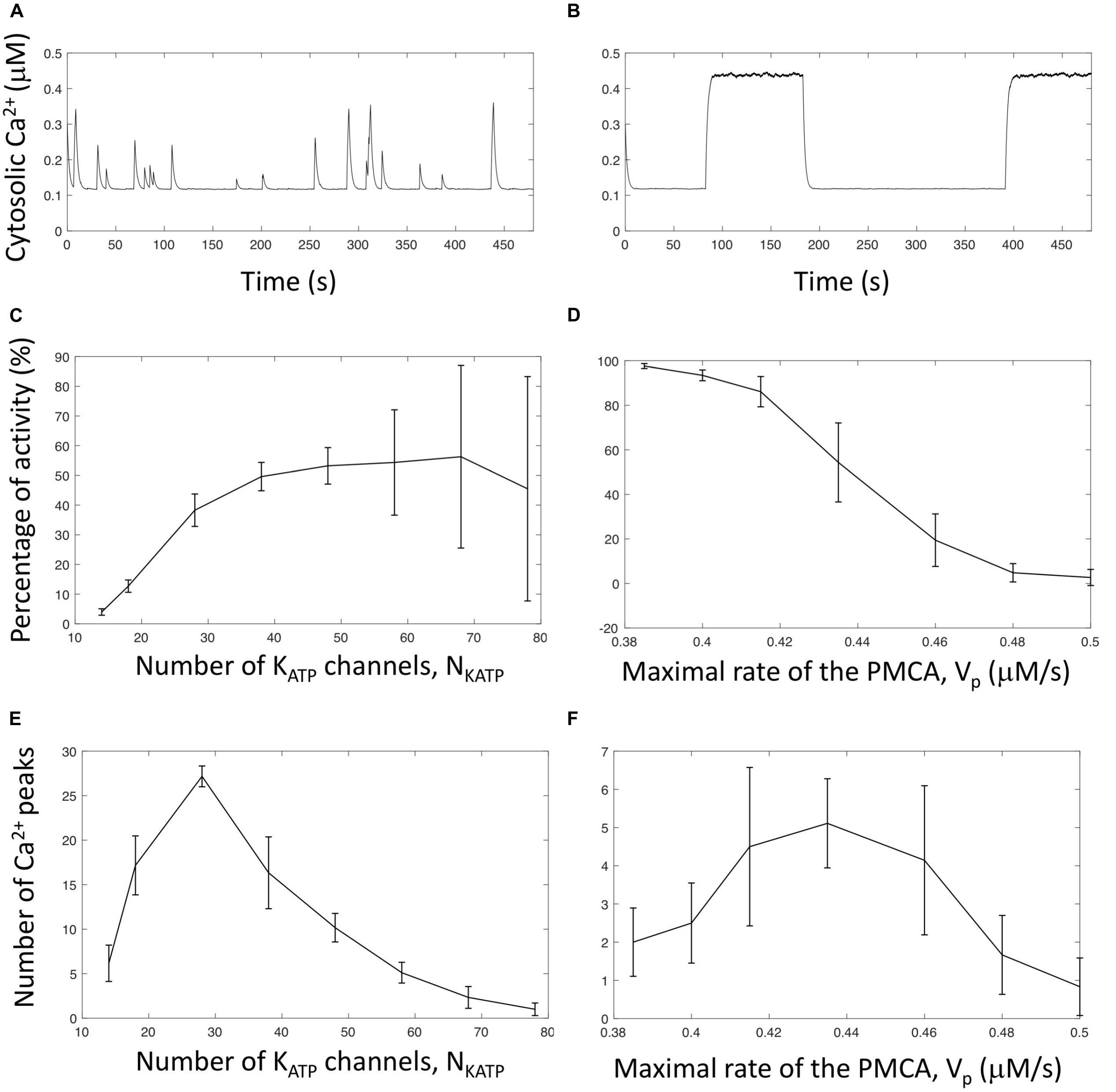


FIGURE 5 | Effect of the number of K_{ATP} channels and rate of PMCA on the pattern of Ca^{2+} oscillations in α-cells. **(A)** Time series of cytosolic Ca^{2+} for $N_{KATP} = 14$ and $g^{s}_{KATP} = 0.170$ nS. The maximal global cellular K_{ATP} conductance ($g^{s}_{KATP} \cdot N_{KATP}$) is the same as in **Figure 4**, but the number of participating channels is smaller. **(B)** Time series of cytosolic Ca^{2+} for $N_{KATP} = 68$ and $g^{s}_{KATP} = 0.03497$ nS. The maximal global cellular K_{ATP} conductance ($g^{s}_{KATP} \cdot N_{KATP}$) is the same as in **Figure 4**, but the number of participating channels is larger. **(C)** Normalized duration of activity as a function of N_{KATP}. **(D)** Normalized duration of activity as a function of the maximal velocity of the PMCA (V_p). **(E)** Number of cytosolic Ca^{2+} peaks in a 500 s period as a function of N_{KATP}. **(F)** Number of cytosolic Ca^{2+} peaks in a 500 s period as a function of the maximal velocity of the PMCA (V_p). Panels **(C,D)** show the times during which cytosolic Ca^{2+} exceeds 0.3 μM, divided by the total simulation time (500 s). These ratios are expressed in percentages. Six independent simulations have been run for $N_{KATP} \leq 58$, and 9 for $N_{KATP} > 58$. For panels **(E,F)**, a peak is counted when Ca_c crosses 0.3 μM with a positive slope. Six independent simulations have been run for all values of V_p.

values of N_{KATP} were accompanied by changes in the single channel unitary conductance (g^{s}_{KATP}) in such a way that the product $N_{KATP} \cdot g^{s}_{KATP}$ remains constant. Given this constraint, the global cell K_{ATP} conductance remains the same for all simulations, which allowed us to only investigate the effect of changes in randomness.

As visible in **Figure 5A**, when the number of channels is very small ($N_{KATP} = 14$), the cell membrane is most of the time hyperpolarized and infrequent, short-duration and low-amplitudes Ca^{2+} spikes occur. If N_{KATP} is smaller than 10, the cell is always in a resting state, although the average value of the cell K_{ATP} conductance (0.21 nS) is in the oscillatory domain (see **Figure 2D**). Fluctuations in g_{KATP} are indeed so important that random depolarizations of the cell membrane are never long enough to initiate an AP. From $N_{KATP} = 10$, infrequent Ca^{2+} spikes start to occur. They are very short in duration and amplitude because, once initiated, they are rapidly aborted by a fluctuation that repolarizes the cell. As a consequence, average Ca^{2+} and ATP concentrations are near their resting levels. These concentrations respectively increase and decrease when considering more channels in the simulations, i.e., when randomness is less pronounced. This is due to an increase in both the number of Ca^{2+} peaks (i.e., in the number of bursts of electrical activity, **Figure 5E**) and the duration of the peaks. Both changes are due to a possibly longer effect of random changes in g_{KATP} allowing to initiate changes in electrical activity. When N_{KATP} gets still larger, simulations exhibit a small number of Ca^{2+} spikes of very long duration (**Figure 5B**). Fluctuations indeed decrease and once in a state, active or resting, the cell tends to remain in this state. In consequence, the number of spikes on a 500 s period becomes smaller (**Figure 5E**), and the average time of activity larger (**Figure 5C**). It should be emphasized that for all simulations presented in **Figures 5C** and **E**, the average value of g_{KATP} remains approximatively the same (∼0.21 nS). In the deterministic analysis, this value corresponds to repetitive AP's (**Figure 2D**). The different behaviors observed for different values of N_{KATP} thus exclusively rely on the characteristics of the noise. This computational observation provides an explanation for the widely different profiles of Ca^{2+} oscillations observed in single α-cells that probably express different number of K$_{ATP}$ channels.

In a given cell characterized by a fixed number of K$_{ATP}$ channels, bursts of electrical activity and Ca^{2+} oscillations are sensitive to factors affecting Ca^{2+} dynamics and particularly to the maximal rate of the PMCA, as shown in **Figures 5D** and **F**. This velocity indeed affects both Ca^{2+} pumping and ATP hydrolysis. When pumping is slower that in the control situation (**Figure 4**), Ca^{2+} and ATP concentrations at the membrane remain large, which tends to decrease the number of bursts of activity and increase their duration. Upon an increase in V_p, bursts become shorter. Thus, the number of spikes in a given period of time reaches a maximum value. When Vp further increases, the decrease in ATP concentration prevents this nucleotide from inducing trends in g_{KATP} changes and Ca^{2+} spikes finally disappear.

Simulations thus predict that the interplay between Ca^{2+} and ATP dynamics that occur through PMCA activity can induce a trend in the noise-initiated changes in K$_{ATP}$ conductance (**Figure 6**). Rapid fluctuations in the cell K$_{ATP}$ conductance occur because of the small number of such channels involved in the electrical activity of α-cells. The resulting changes in membrane Ca^{2+} induce changes in ATP concentration, because Ca^{2+} and ATP levels in the submembrane space are coupled via the activity of the PMCA. ATP changes in turn feedback on the conductance of the K$_{ATP}$ channels. As the concentration of this nucleotide evolves slowly, these changes are in average maintained on a period of time that is sufficiently long to induce or refrain electrical activity. Thus, the combination of randomness at the level of the K$_{ATP}$ current and of ATP changes can account for the observed long-duration changes in electrical activity and thus for the Ca^{2+} oscillations observed in α-cells.

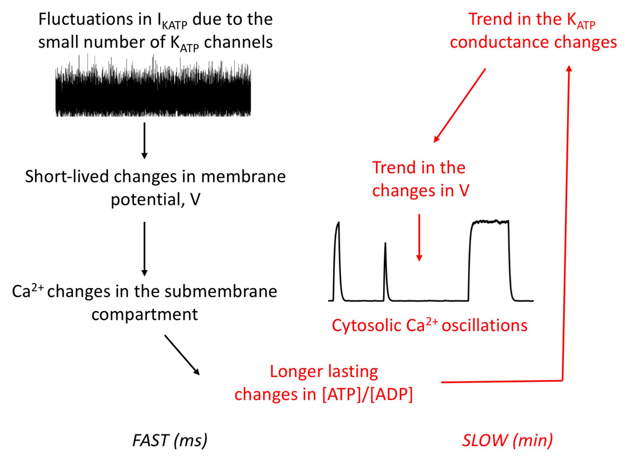


FIGURE 6 | Schematic representation of the interplay between randomness of the K$_{ATP}$ current and ATP changes to generate irregular, slow Ca^{2+} oscillations in α-cells. Rapid fluctuations in the K$_{ATP}$ current due to the small number of channels in the α-cell plasma membrane create fluctuations in membrane voltage that can induce a short-lived change in membrane voltage, which in turn modifies submembrane Ca^{2+} concentration. If the dynamics of ATP is not considered, this noise-induced change in Ca^{2+} would last some milliseconds. If ATP is considered, large fluctuations create a change in ATP concentration of larger duration, because ATP kinetics is much slower than electrical activity. Despite the rapid noise on V that is still present, this will create a trend in the changes in V because of the feedback exerted by ATP on the K$_{ATP}$ current. Thus, periods of quiescence and activity will alternate, which is responsible for cytosolic Ca^{2+} oscillations by diffusion of the membrane Ca^{2+}.

DISCUSSION

In low glucose conditions, more than half of α-cells display Ca^{2+} oscillations and these oscillations can persist in high glucose, although much reduced in frequency and amplitude. These Ca^{2+} oscillations are highly variable in shape, duration and interspike interval, not only among different cells but also in the course of time in a given individual cell. Such a randomness strongly suggests the presence of a high level of molecular noise, which generally results from a low number of copies of one or several biochemical species (Gonze et al., 2018). In direct agreement with this theoretical concept, α-cells electrical activity involves a limited number of K$_{ATP}$ channels that play a crucial role in their electrical activity (Rorsman et al., 2014). The aim of this study was to investigate the link between the small numbers of K$_{ATP}$ channels and the existence of irregular, slow cytosolic Ca^{2+} oscillations in α-cells in the rigorous framework of a mathematical model closely based on experimental observations. As Ca^{2+} oscillations are observed not only in α-cells of the intact pancreas but also in isolated cells (Salehi et al., 2006; Quoix et al., 2009; Tuduri et al., 2009; Le Marchand and Piston, 2010), they most probably result from an intrinsic mechanism, modulated by paracrine signaling.

We thus built a model of increasing complexity, on the basis of a previously proposed model of the α-cells electrical activity (Diderichsen and Göpel, 2006). This sequential approach showed that the randomness of the K_{ATP} current does not *per se* allow for Ca^{2+} oscillations. Random changes in this current are indeed too fast to induce any long-term change that could induce intermittent electrical activity. However, when taking into account the triangular relationship between electrical current, Ca^{2+} changes and ATP consumption, the model can account for the electrical and calcium dynamics experimentally observed in α-cells. Moreover, it also accounts for the fact that these dynamics are not stereotypic as quite moderate changes in the number of K_{ATP} channels can induce significant changes in the pattern of Ca^{2+} oscillations. The model predicts that factors that interfere with cell Ca^{2+} changes – such as pumping rates, diffusion coefficients or buffering capacities – may modify the durations during which repetitive AP's occur. This prediction is relevant for the impaired glucagon response to hypoglycemia and hyperglucagonemia observed in type 2 diabetes.

The simulations predict that the number of simultaneously active K_{ATP} channels during electrical activity is between ∼5 and 10, in agreement with experimental observations (Zhang et al., 2020). This number is much smaller than the total number of channels considered in the simulations ($N_{KATP} = 58$), which are all potentially openable. Their intrinsically low open probability and their inhibition by ATP explain why only a small fraction are open simultaneously. Moreover, we expect our value of N_{KATP} to be underestimated since we did not take into account effective fatty acid metabolism in low glucose conditions (Briant et al., 2018a) nor the spatio-temporal dynamics of ATP. In any case, that intermittency in electrical activity can be obtained with the model over a relatively large range of values of N_{KATP}, as shown in **Figure 5**, indicates that the proposed mechanism for Ca^{2+} oscillations in α-cells is robust.

A similar role for stochasticity at the level of the K_{ATP} current has been put forward to account for the irregularity of the neuronal firing pattern in hippocampal CA3 neurons (Huang et al., 2007). However, in this case, the changes in K_{ATP} current are due to rapid, random fluctuations in the concentration of ATP, which results in an irregular frequency of AP's and not in intermittent electrical activity. In the more general context of networks dynamics, it is known that fluctuations can propagate along the different nodes of the network with a rate of decay or enhancement that depends on the network's structure (Maslov et al., 2007; Zhang et al., 2012).

Within this study, we did not investigate the behavior of the model when changing external glucose or considering paracrine signaling, which are left for further study. These limitations are due to the fact that the model used for electrical activity (Diderichsen and Göpel, 2006) does not reproduce one of the key characteristics of α-cells when raising external glucose, i.e., the reduction in the amplitude of the AP's (see Ashcroft and Rorsman, 2013 for example). This reduction leads to less activation of the Ca^{2+} channels linked to glucagon secretion. In **Figure 2**, this reduction was simulated by an artificial change in the leak conductance when changing the conductance of the K_{ATP} channels to simulate the changes in external glucose. However, given the reliability of the Diderichsen and Göpel's model and the many unknowns in the field of α-cells dynamics, we considered this unchanged model as a safe starting point to explore the impact of a stochastic K_{ATP} current and of changes in ATP concentration. However, we acknowledge that values of parameters for the ionic currents obtained with better methods of α-cell identification could improve the agreement between the simulations and the experiments (Briant et al., 2017; Montefusco et al., 2020).

The hydrolysis of ATP that parallels Ca^{2+} extrusion out of the cell plays a key role in the behavior predicted by the model. However, ATP synthesis is also sensitive to intracellular Ca^{2+} changes because mitochondrial metabolism is Ca^{2+}-sensitive (Wacquier et al., 2019). The ATP/ADP ratio is also controlled by glycolysis and fatty acid metabolism (Briant et al., 2018a). Further studies are required to address these interrelationships in a more accurate way (Olivos-Santes et al., 2019), although the qualitative results obtained with the relatively simple model used in this study should be robust towards these numerous possible model refinements.

As another limitation, we only considered noise at the level of the K_{ATP} channel, while all molecular processes, and especially Ca^{2+} dynamics, are subjected to noise. This simplification was based on the experimental observation of an unusually small K_{ATP} conductance in α-cells, in particular in comparison with pancreatic β-cells that are known to be equipped with the same channels but still display very regular voltage and calcium patterns. However, the details of our conclusions may be affected by adopting a full stochastic description of the cell dynamics. In particular, it may lead to less stereotypic cytosolic Ca^{2+} changes, as reported in the experiments.

An interesting perspective would be to couple the model presented in this study to our previously proposed model that relates glucagon secretion to cytosolic Ca^{2+} changes (González-Vélez et al., 2012). The latter model is able to predict glucagon secretion using as an input experimentally obtained Ca^{2+} traces. It was found that glucagon secretion does not correlate with the frequency of Ca^{2+} oscillations. This investigation could be pursued by using Ca^{2+} time series obtained by simulations of the present model instead of experimental data. This could lead to identify the key elements of electrical and/or calcium activities controlling glucagon secretion.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

VG-V, AP, and GD contributed to the conceptualization, design of the study, development, simulation, and analysis of the models. GD and VG-V contributed to the preparation of the manuscript. All authors contributed to the article and approved the final version.

FUNDING

GD is research director at the Belgian FRS-FNRS. VG-V and GD acknowledge previous financial support of FNRS and CONACYT México through a bilateral cooperation project.

REFERENCES

Allbritton, N., Meyer, T., and Stryer, L. (1992). Range of messenger action of calcium ion and inositol 1,4,5-trisphosphate. *Science* 258, 1812–1815. doi: 10.1126/science.1465619

Ashcroft, F., and Rorsman, P. (2013). KATP channels and islet hormone secretion: new insights and controversies. *Nat. Rev. Endocrinol.* 9, 660–669. doi: 10.1038/nrendo.2013.166

Barg, S., Galvanovskis, J., Göpel, S., Rorsman, P., and Eliasson, L. (2000). Tight coupling between electrical activity and exocytosis in mouse glucagon-secreting α-cells. *Diabetes Metab. Res. Rev.* 49, 1500–1510. doi: 10.2337/diabetes.49.9.1500

Bokvist, K., Olsen, H., Hoy, M., Gotfredsen, C., Holmes, W., Buschard, K., et al. (1999). Characterisation of sulphonylurea and ATP-regulated Ca2+ channels in rat pancreatic α-cells. *Eur. J. Physiol.* 438, 428–436. doi: 10.1007/s004240051058

Briant, L., Dodd, M., Chibalina, M., Rorsman, N., Johnson, P., Carmeliet, P., et al. (2018a). CPT1a-dependent long-chain fatty acid oxidation contributes to maintaining glucagon secretion from pancreatic islets. *Cell Rep.* 23, 3300–3311. doi: 10.1016/j.celrep.2018.05.035

Briant, L., Reinbothe, T., Spiliotis, I., Miranda, C., Rodriguez, B., and Rorsman, P. (2018b). δ-cells and β-cells are electrically coupled and regulate α-cell activity via somatostatin. *J. Physiol.* 596, 197–215.

Briant, L., Salehi, A., Vergari, E., Zhang, Q., and Rorsman, P. (2016). Glucagon secretion from pancreatic α-cells. *Upsala J. Med. Sci.* 121, 113–119. doi: 10.3109/03009734.2016.1156789

Briant, L., Zhang, Q., Vergari, E., Kellard, J., Rodriguez, B., Ashcroft, F., et al. (2017). Functional identification of islet cell types by electrophysiological fingerprinting. *J. R. Soc. Interface* 14:20160999. doi: 10.1098/rsif.2016.0999

Csercsik, D., Farkas, I., Szederkenyi, G., Hrabovszky, E., Liposits, Z., and Hangos, K. (2010). Hodgkin-Huxley type modelling and parameter estimation of GnRH neurons. *Biosystems* 100, 198–207. doi: 10.1016/j.biosystems.2010.03.004

D'Alessio, D. (2011). The role of dysregulated glucagon secretion in type 2 diabetes. *Diabetes. Obes. Metab.* 13(Suppl. 1), 126–132. doi: 10.1111/j.1463-1326.2011.01449.x

Diderichsen, P., and Göpel, S. (2006). Modelling the electrical activity of pancreatic α-cells based on experimental data from intact mouse islets. *J. Biol. Phys.* 32, 209–229. doi: 10.1007/s10867-006-9013-0

Dupont, G., Falcke, M., Kirk, V., and Sneyd, J. (2016). *Models of Calcium Signalling*. Berlin: Springer International Publishing., doi: 10.1007/978-3-319-29647-0

Enkvetchakul, D., Loussouam, G., Makhina, E., and Nichols, C. (2001). ATP interaction with the open state of the KATP channel. *Biophys. J.* 80, 719–728. doi: 10.1016/s0006-3495(01)76051-1

Ermentrout, B. (2002). *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. Philadelphia, PA: SIAM.

Fall, C., Marland, E., Wagner, J., and Tyson, J. (2005). *Computational Cell Biology*. New York, NY: Springer.

Fridlyand, L., and Philipson, L. (2012). A computational systems analysis of factors regulating a cell glucagon secretion. *Islets* 4, 262–283. doi: 10.4161/isl.22193

Gilon, P. (2020). The role of α-cells in islet function and glucose homeostasis in health and type 2 diabetes. *J. Mol. Biol.* 432, 1367–1394. doi: 10.1016/j.jmb.2020.01.004

González-Vélez, V., Dupont, G., Gil, A., Gonzalez, A., and Quesada, I. (2012). Model for glucagon secretion by pancreatic α-cells. *PLoS One* 7:e32282. doi: 10.1371/journal.pone.0032282

ACKNOWLEDGMENTS

We thank Vincent Verjans for scientific advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.602844/full#supplementary-material

González-Vélez, V., Gil, A., and Quesada, I. (2010). Minimal state models for ionic channels involved in glucagon secretion. *Math. Biosci. Eng.* 7, 793–807. doi: 10.3934/mbe.2010.7.793

Gonze, D., Gérard, C., Wacquier, B., Woller, A., Tosenberg, A., and Dupont, G. (2018). Modeling-based investigation of the effect of noise in cellular systems. *Front. Mol. Biosc.* 5:34. doi: 10.3389/fmolb.2018.00034

Göpel, S., Zhang, Q., Eliasson, L., Ma, X. S., Galvanovskis, J., and Kanno, T. (2004). Capacitance measurements of exocytosis in mouse pancreatic α-, β- and δ-cells within intact islets of Langerhans. *J. Physiol.* 556, 711–726. doi: 10.113/jphysiol.2003.059675

Grubelnik, V., Markovic, R., Lipovsek, S., Leitinger, G., Gosak, M., Dolensek, J., et al. (2020). Modelling of dysregulated glucagon secretion in type 2 diabetes by considering mitochondrial alterations in pancreatic α-cells. *R. Soc. Open Sci.* 7:191171. doi: 10.1098/rsos.191171

Huang, C. W., Huang, C. W., Chen, J., Tsai, J., and Wu, S. (2007). Glucose and hippocampal neuronal excitability: role of ATP-sensitive potassium channels. *J. Neurosci. Res.* 85, 1468–1477. doi: 10.1002/jnr.21284

Huang, Y.-C., Gaisano, H., and Leung, Y.-M. (2011). Electrophysiological identification of mouse islet -cells. From isolated single α-cells to in situ assessment within pancreas slices. *Islets* 3, 139–143. doi: 10.4161/isl.3.4.16166

Kellard, A., Rorsman, N., Hill, T., Armour, S., van de Bunt, M., Rorsman, P., et al. (2020). Reduced somatostatin signalling leads to hypersecretion of glucagon in mice fed a high-fat diet. *Mol. Metab.* 40:101021. doi: 10.1016/j.molmet.2020.101021

Khan, F., Goforth, P., Zhang, M., and Satin, L. (2001). Insulin activates ATP-sensitive K+ channels in pancreatic β-cells through a phosphatidyl-inositol 3-kinase-dependent pathway. *Diabetes Metab. Res. Rev.* 50, 2192–2198. doi: 10.2337/diabetes.50.10.2192

Le Marchand, S., and Piston, D. (2010). Glucose suppression of glucagon secretion. *J. Biol. Chem.* 285, 14389–14398.

Le Marchand, S., and Piston, D. (2012). Glucose decouples intracellular Ca2+ activity from glucagon secretion in mouse pancreatic islet alphα-cells. *PLoS One* 7:e47084. doi: 10.1371/journal.pone.0047084

Li, J., Yu, Q., Ahooghalandari, P., Gribble, F., Reimann, F., Tengholm, A., et al. (2015). Submembrane ATP and Ca2+ kinetics in α-cells: unexpected signaling for glucagon secretion. *FASEB J.* 29, 3379–3388. doi: 10.1096/fj.14-265918

MacDonald, P., De Marinis, Y. Z., Ramracheya, R., Salehi, A., Ma, X., Johnson, P., et al. (2007). A KATP channel-dependent pathway within α cells regulates glucagon release from both rodent and human islets of Langerhans. *PLoS Biol.* 5:e143. doi: 10.1371/journal.pbio.0050143

Maslov, S., Sneppen, K., and Ispolatov, I. (2007). Spreading out of perturbations in reversible reaction networks. *New J. Phys.* 9:273. doi: 10.1088/1367-2630/9/8/273

Montefusco, F., Cortese, G., and Pedersen, M. (2020). Heterogeneous alphα-cell population modeling of glucose-induced inhibition of electrical activity. *J. Theor. Biol.* 485:110036. doi: 10.1016/j.jtbi.2019.110036

Montefusco, F., and Pedersen, M. G. (2015). Mathematical modelling of local calcium and regulated exocytosis during inhibition and stimulation of glucagon secretion from pancreatic alphα-cells. *J. Physiol.* 593, 4519–4530. doi: 10.1113/jp270777

Olivos-Santes, E., Romero-Campos, H., Dupont, G., and González-Vélez, V. (2019). A modelling study of glycolytic oscillations and electrical activity in pancreatic alpha cells. *Rev. Mex. De Ing. Biom.* 40, 1–10.

Quesada, I., Tuduri, E., Ripoll, C., and Nadal, A. (2008). Physiology of the pancreatic α-cell and glucagon secretion: role in glucose homeostasis and diabetes. *J. Endocrin.* 199, 5–19. doi: 10.1677/joe-08-0290

Quoix, N., Cheng-Xue, R., Mattart, L., Zeinoun, Z., Guiot, Y., Beauvois, M., et al. (2009). Glucose and pharmacological modulators of ATP-sensitive K+ channels control [Ca2+]c by different mechanisms in isolated mouse α-cells. *Diabetes Metab. Res. Rev.* 58, 412–421. doi: 10.2337/db07-1298

Rorsman, P., Ramracheya, R., Rorsam, N., and Zhang, Q. (2014). ATP-regulated potassium channels in pancreatic alpha and beta cells: similar functions but reciprocal effects on secretion. *Diabetologia* 57, 1749–1761. doi: 10.1007/s00125-014-3279-8

Salehi, A., Vieir, E., and Gylfe, E. (2006). Paradoxical stimulation of glucagon secretion by high glucose concentrations. *Diabetes Metab. Res. Rev.* 55, 2318–2323. doi: 10.2337/db06-0080

Tuduri, E., Marroqui, L., Soriano, S., Ropero, A., Batista, T., Piquer, S., et al. (2009). Inhibitory effects of leptin on pancreatic α-cell function. *Diabetes Metab. Res. Rev.* 58, 1616–1624. doi: 10.2337/db08-1787

Voets, T. (2000). Dissection of three Ca2+-dependent steps leading to secretion in chromaffin cells from mouse adrenal slice. *Neuron* 28, 537–545. doi: 10.1016/s0896-6273(00)00131-8

Wacquier, B., Combettes, L., and Dupont, G. (2019). Cytoplasmic and mitochondrial calcium signaling: a two-way relationship. *Cold Spring Harb. Perspect. Biol.* 11:a035139. doi: 10.1101/cshperspect.a035139

Watts, M., Ha, J., Kimchi, O., and Sherman, A. (2016). Paracrine regulation of glucagon secretion: the β/α/δ model. *Am. J. Physiol. Endocrinol. Metab.* 310, E597–E611.

Watts, M., and Sherman, A. (2014). Modeling the pancreatic α-cell: dual mechanisms of glucose suppression of glucagon secretion. *Biophys. J.* 106, 741–751. doi: 10.1016/j.bpj.2013.11.4504

Wendt, A., and Eliasson, L. (2020). Pancreatic -cells – The unsung heroes in islet function. *Sem. Cell and Dev. Biol.* 103, 41–50. doi: 10.1016/j.semcdb.2020.01.006

Yu, A., Shuai, H., Ahooghalandari, P., Gylfe, E., and Tengholm, A. (2019). Glucose controls glucagon secretion by directly modulating cAMP in alpha cells. *Diabetologia* 61, 1212–1224. doi: 10.1007/s00125-019-4857-6

Zhang, H., Chen, Y., and Chen, Y. (2012). Noise propagation in gene regulation networks involving interlinked positive and negative feedback loops. *PLoS One* 7:e51840. doi: 10.1371/journal.pone.0051840

Zhang, Q., Dou, H., and Rorsman, P. (2020). 'Resistance is futile?'– paradoxical inhibitory effects of KATP channel closure in glucagon-secreting α-cells. *J. Physiol.* 598, 4765–4780. doi: 10.1113/JP279775

Zhang, Q., Ramracheya, R., Lahmann, C., Tarasov, A., Bengtsson, M., Braha, O., et al. (2013). Role of KATP channels in glucose-regulated glucagon secretion and impaired counterregulation in Type 2 Diabetes. *Cell Metab.* 18, 871–882. doi: 10.1016/j.cmet.2013.10.014

Zimny, M., and Blackard, W. (1975). The surface structure of isolated pancreatic islet cells. *Cell Tissue Res.* 164, 467–471.

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

frontiers
in Physiology

ORIGINAL RESEARCH
published: 30 November 2020
doi: 10.3389/fphys.2020.579117

Mechanistic Model of Signaling Dynamics Across an Epithelial Mesenchymal Transition

James D. Wade[1,2†], **Xiao-Kang Lun**[3†], **Nevena Zivanovic**[3], **Eberhard O. Voit**[1*‡] and **Bernd Bodenmiller**[2*]

[1] Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Emory University, Atlanta, GA, United States, [2] Department of Quantitative Biomedicine, University of Zürich, Zürich, Switzerland, [3] Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland

Edited by:
Zhike Zi,
Max Planck Institute for Molecular
Genetics, Germany

Reviewed by:
Kwang-Hyun Cho,
Korea Advanced Institute of Science
and Technology, South Korea
Xiaojun Tian,
Arizona State University, United States
Xuefei Li,
Shenzhen Institutes of Advanced
Technology, China

***Correspondence:**
Bernd Bodenmiller
bernd.bodenmiller@uzh.ch
Eberhard O. Voit
eberhard.voit@bme.gatech.edu

†Present address:
James D. Wade,
LYO-X GmbH, Allschwil, Switzerland
Xiao-Kang Lun,
Wyss Institute for Biologically Inspired
Engineering, Harvard University,
Boston, MA, United States

‡Lead contact

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: 01 July 2020
Accepted: 16 October 2020
Published: 30 November 2020

Citation:
Wade JD, Lun X-K, Zivanovic N,
Voit EO and Bodenmiller B (2020)
Mechanistic Model of Signaling
Dynamics Across an Epithelial
Mesenchymal Transition.
Front. Physiol. 11:579117.
doi: 10.3389/fphys.2020.579117

Intracellular signaling pathways are at the core of cellular information processing. The states of these pathways and their inputs determine signaling dynamics and drive cell function. Within a cancerous tumor, many combinations of cell states and microenvironments can lead to dramatic variations in responses to treatment. Network rewiring has been thought to underlie these context-dependent differences in signaling; however, from a biochemical standpoint, rewiring of signaling networks should not be a prerequisite for heterogeneity in responses to stimuli. Here we address this conundrum by analyzing an *in vitro* model of the epithelial mesenchymal transition (EMT), a biological program implicated in increased tumor invasiveness, heterogeneity, and drug resistance. We used mass cytometry to measure EGF signaling dynamics in the ERK and AKT signaling pathways before and after induction of EMT in Py2T murine breast cancer cells. Analysis of the data with standard network inference methods suggested EMT-dependent network rewiring. In contrast, use of a modeling approach that adequately accounts for single-cell variation demonstrated that a single reaction-based pathway model with constant structure and near-constant parameters is sufficient to represent differences in EGF signaling across EMT. This result indicates that rewiring of the signaling network is not necessary for heterogeneous responses to a signal and that unifying reaction-based models should be employed for characterization of signaling in heterogeneous environments, such as cancer.

Keywords: epithelial-to-mesenchymal transition (EMT), systems biology, ordinary differential equations, kinetic model, ERK pathway, AKT pathway, computational biology, single cell modeling

1. INTRODUCTION

Intracellular signaling networks are biochemical systems that integrate spatio-temporal information regarding the intra- and extracellular environments into functional programs that drive cellular decisions (Dolmetsch et al., 1997; Kholodenko, 2006; Selimkhanov et al., 2014). Signals, such as extracellular ligand concentrations, are transduced by modulating the enzymatic activities and local concentrations of signaling mediators, such as kinases, within a cell. In traditional analysis, the structures of signaling networks are studied in biochemical experiments and are then formalized as graphs where nodes represent the activity state of signaling molecules and directed edges represent interactions between molecules. More recently, statistical modeling has been used to infer network structures in a data-driven manner (Sachs et al., 2005). In contrast to canonical representations of signaling networks as static structures, data-driven approaches of network inference have represented the structure of signaling networks as dependent upon context, including cell phenotype, type of input signal, and treatment, e.g., with an inhibitor, as well as unaccounted cell-to-cell variability (Hill et al., 2017; Petsalaki et al., 2015; Will and Helms, 2015; Brightman and Fell, 2000). The context-dependence of signaling is of particular consequence in diseases like cancer, where genetic errors lead to changes in the relative expression or function (Creixell et al., 2015) of signaling proteins and where the local microenvironment can be abnormal, resulting in signal responses that substantially differ from those of healthy cells (Altschuler and Wu, 2010). Consideration of signaling network rewiring between contexts has led to the development of novel treatment regimens (Lee et al., 2012). While useful, data-driven network inference requires large quantities of data and, typically, some prior knowledge of causative relationships. However, even though contexts are considered in this type of modeling, the results remain static representations of dynamic processes, which immediately limits predictions of cellular dynamics and responses (Kolitz and Lauffenburger, 2012).

Unlike graph-based network models, mechanistic models of signaling capture both the reaction network structure and the temporal dynamics of signaling, with residual noise attributed to natural cell-to-cell variability or stochasticity. These models consist of sets of reactions in the form of differential equations that describe how each signaling component changes over time as a function of the overall state of the system, and all model components and parameters have physical interpretations, such as concentrations, binding affinities, or reaction rates (Aldridge et al., 2006). Properly calibrated mechanistic models can be used to predict cellular dynamics from a snapshot of cell state and to analyze the consequences of observed or hypothetical alterations in relative concentrations or activities of components. The drawback of mechanistic models is that their construction relies on detailed prior knowledge of the reaction network structure and on multiple, targeted experiments that permit calibration of the model's parameter values (Aldridge et al., 2006; Kholodenko, 2006). The need for detailed prior information presents a great challenge when multiple contexts are to be considered (Halasz et al., 2016; Kholodenko, 2006).

Biochemical reasoning suggests that the signaling reaction structures themselves should be fixed, as the kinetics of a particular interaction does not change unless other factors or modulators are altered. This reasoning suggests that a mechanistic model should be "context-explicit," reconciling the entire range of context-dependent signaling dynamics without network rewiring. In principle, therefore, it should be possible to construct a single mechanistic model that, combined with initial snapshots of individual cell states, explains and predicts signaling across many phenotypic contexts. In practice, however, the ability of a single mechanistic model to represent signaling across cell phenotypes is often limited due to ill-characterized differences in cellular milieu and gaps in knowledge, which are often considered as sources of natural stochasticity.

Here, we propose differential equation based mechanistic modeling—a model with static network structure—of single-cell mass cytometry data to investigate whether a context-dependent signaling network structure is necessary to represent differences in signaling dynamics in two distinct cell phenotypes: cells before and after an epithelial-mesenchymal transition (EMT). EMT is a developmental program through which epithelial cells trans-differentiate into a mesenchymal phenotype; EMT results in loss of cell-cell adhesion junctions, increased capacity for migration and invasion, and resistance to apoptosis (Fu et al., 2018). EMT has been implicated in the generation of metastatic and resistant cancer cell populations, and both bulk (Desai et al., 2015) and single-cell data (Krishnaswamy et al., 2018) have demonstrated EMT-associated alterations in signaling.

To study changes in signaling dynamics and network structure across an EMT, we use a previously established experimental model of TGF-β-induced EMT in Py2T murine breast cancer cells (Waldmeier et al., 2012). We performed mass cytometry to simultaneously quantify markers of epithelial and mesenchymal phenotypes, as well as the total expression and phosphorylation dynamics of multiple MAPK/ERK and PI3K/AKT signaling pathway components in response to stimulation with EGF, a proliferative signal that can drive tumor growth. These pathways and EGF signaling were chosen due to the robust signaling dynamics they provided in both epithelial and mesenchymal Py2T cell phenotypes.

To study the EMT-related changes in EGF signaling, we first applied a classical network-inference method to the data. This analysis suggested that the network structure of the ERK and AKT pathways is rewired in a phenotype-dependent manner. Then, using the same data and a mechanistic single-cell modeling approach based on the core principles of biochemical systems modeling (Savageau, 1976; Voit, 2000, 2013), we constructed, *ab initio*, two models of the signaling pathway, one for the mesenchymal phenotype and one for the epithelial phenotype. Intriguingly, accounting for unmeasured contextual variables allowed us to consolidate the two models into a single model with constant reaction structure and with very modest residual noise. This result presents proof-of-principle that rewiring does not occur during EMT but that alterations in signaling processes during this dramatic transition result from changes in the relative concentrations of signaling components. Our results suggest that further extending this type of single-cell model development, combined with highly-multiplexed single-cell measurements of intra- and extracellular states, has the potential to greatly reduce uncertainty that is otherwise attributed to stochasticity and to improve our ability to predict and analyze signaling responses in heterogeneous tissues and different disease contexts.

2. RESULTS

During EMT, cells undergo dramatic changes in phenotype that are known to be driven by signaling; therefore, we sought to determine whether signaling networks are rewired or whether changes in the relative concentrations of signaling components are sufficient to permit altered dynamics (**Figure 1**). Specifically,

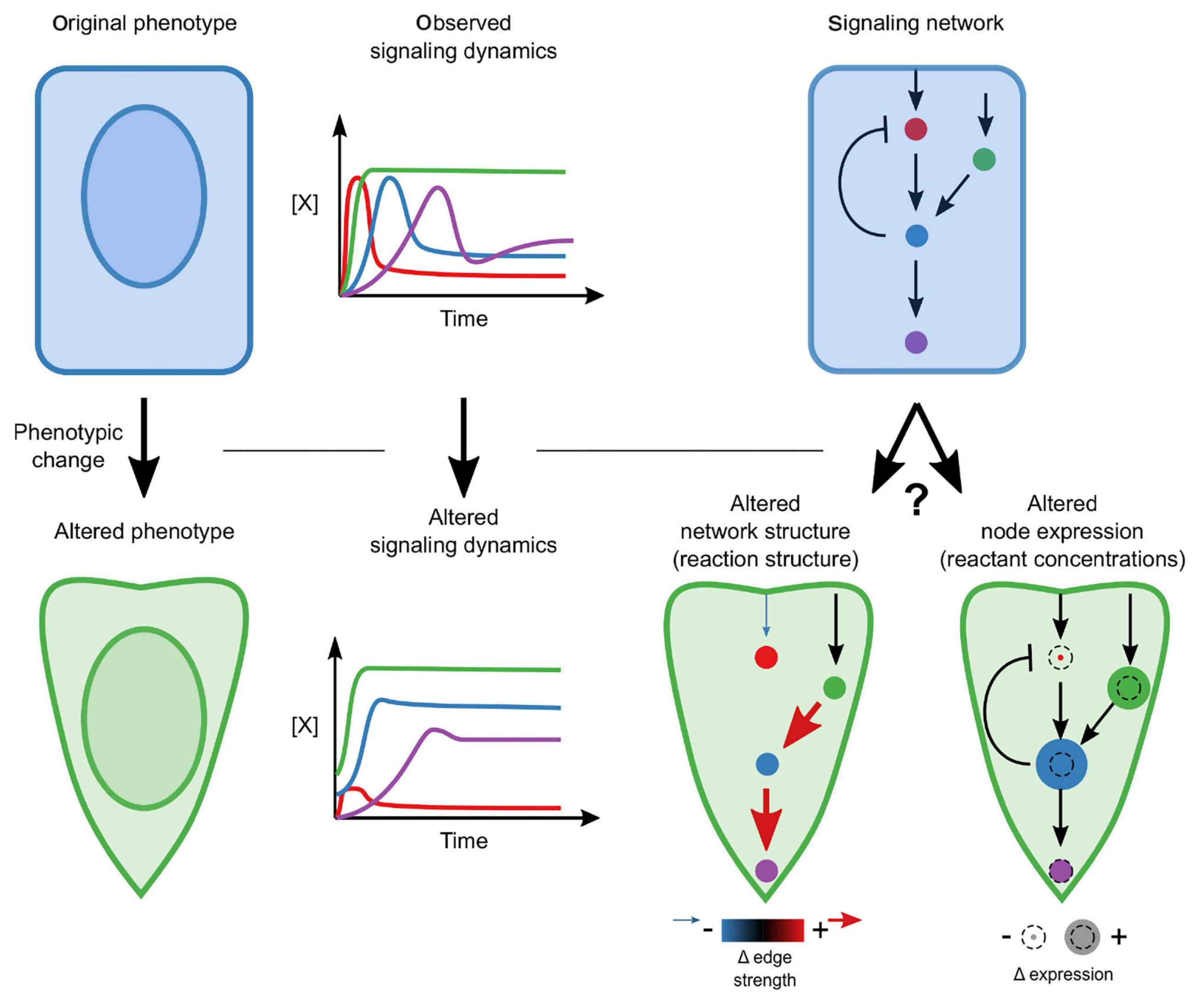


FIGURE 1 | Conceptual overview. A phenotypic transition is associated with a changed signaling response. A key question is whether this alteration requires network rewiring (associated with the addition or elimination of edges) or whether the network structure is constant and moderate differences in relative concentrations drive differences.

we pose the hypothesis that a single, mechanistic model with fixed reaction network structure and kinetic parameters, calibrated with highly-multiplexed single-cell measurements of signaling protein state and expression, should be able to reconcile the differences in signaling across phenotypes.

To test our hypothesis, we used an experimental model of EMT, TGF-β-induced EMT in murine Py2T cells, and focused on EGF signaling dynamics in the proliferative, pro-growth, and pro-oncogenic MAPK/ERK and PI3K/AKT pathways (Wee and Wang, 2017). We used mass cytometry to quantify markers of epithelial and mesenchymal phenotypes and levels of phosphorylated and total signaling proteins. Specifically, we measured total and phosphorylated forms of MEK, ERK, and p90RSK (RSK) in the core ERK pathway, AKT and GSK3β in the AKT pathway, ribosomal protein S6 (S6), a downstream target of both ERK and AKT signaling, as well as cell cycle, death and size markers in 12 serial samples from a time course of EGF stimulation (for the full antibody panel, see **Supplementary Table 4**).

Our dataset contains 31-dimensional measurements across 34 conditions: Thirteen time points for EGF stimulation and four time points for unstimulated EGF control both before and after induction of EMT. Samples were analyzed in triplicate with an average of nearly 10,000 cells analyzed per sample.

2.1. Data-Driven Statistical Network Inference Suggests That the AKT/ERK Signaling Network Is Rewired in Response to TGF-β Treatment

To study signaling during EMT, we used a model of TGF-β induced EMT in Py2T cells (Waldmeier et al., 2012;

Krishnaswamy et al., 2018). When grown in a monolayer, these cells are epithelial in visual and molecular phenotype with a homogeneous cobblestone appearance and high expression of the epithelial marker E-cadherin. Chronic treatment of cells with TGF-β causes them to transition to a mesenchymal phenotype; cells no longer grow in a monolayer, have an elongated shape, no longer express E-cadherin, and instead express mesenchymal markers, such as vimentin. As in previous studies, we defined epithelial cells as those with high levels of E-cadherin and low levels of vimentin expression. To generate a population of mesenchymal cells, samples were treated with TGF-β at 24 h intervals for 3 days; these cells expressed little E-cadherin and high levels of vimentin (**Figure 2A**; **Supplementary Figure 1**).

We first used traditional methods of statistical network inference to determine whether they would suggest that the ERK/AKT signaling network becomes rewired during EMT. Specifically, we used the data-driven approach of partial correlation analysis (Garmaroudi et al., 2010; Desai et al., 2015), which quantifies the correlation between two variables after removing potentially confounding correlations to other variables in the system. The partial correlation between all pairs of phosphoproteins in our panel for epithelial and mesenchymal cells is shown in **Figure 2B** (see **Supplementary Figure 2** for the analysis including both phospho- and total proteins). Partial correlation values with magnitudes >0.1 were taken to define edges representing true interactions. In epithelial cells, partial correlation recovered the canonical pMEK-ppERK-pRSK-pS6 and pAKT-pGSK3β pathways (Mendoza et al., 2011; Manning and Cantley, 2007; Olayioye and Neve, 2000; Wee and Wang, 2017); note that GSK3β phosphorylation at S9 is inhibitory. By contrast, the canonical pAKT-pS6 (via p70S6 kinase) (Mendoza et al., 2011) and pRSK-pGSK3β (Manning and Cantley, 2007) relationships were attributed to pGSK3β-pS6 and ppERK-pGSK3β, respectively. The crosstalk suggested between pMEK and pAKT has not been reported previously, but could indirectly reflect known crosstalk between PI3K3CA and the RAF activator RAC (Ebi et al., 2013).

This analysis suggests that EMT rewired the signaling network. With the chosen cutoff, the mesenchymal cells appear to lose three (of seven) original edges and gain six new edges after the transition to mesenchymal cells. Compared to epithelial cells, in mesenchymal cells the edges ppERK-pRSK, pGSKβ-pS6, or pMEK-pAKT are lost and edges pMEK-pRSK, pMEK-pGSKβ, pMEK-pS6, ppERK-pAKT, ppERK-pS6, and pRSK-pGSKβ are gained. Noticeably absent in mesenchymal cells is a direct path from the AKT pathway to pS6. When different cutoffs were used, the specific edges gained and lost changed somewhat (see **Supplementary Figure 2**), but significant EMT-dependent rewiring was predicted at all cutoffs.

Signaling is a dynamic process. To assess the time dependence of various signaling states, we quantified the dynamics of network relationships by calculating the partial correlation between widely documented "canonical" network edges for each time point (**Figure 2C**). This analysis demonstrated that the strengths of most edge relationships, including that of the pMEK-ppERK edge, varied when compared before and after EMT. Such variation is expected, as existence of an edge can be interpreted as identification of signaling "activation." Notably, at certain time points we observed two edges that were not observed at the steady state, the ppERK-pRSK edge in mesenchymal cells and the pAKT-pS6 edge in epithelial cells. The dependence of the network structure on context, such as time relative to stimulation and cell phenotype, illustrates that such a purely network-based statistical analysis, even if it is based on distributions of single-cell data, does not always accurately predict signaling responses.

2.2. Mechanistic Model With Constant Network Structure Permits Heterogeneity of Epithelial and Mesenchymal ERK/AKT Signaling

We hypothesized that a dynamic mechanistic model with constant reaction network structure that includes context-dependent parameters should more accurately predict signaling dynamics in cells of both epithelial and mesenchymal phenotypes. Using prior knowledge combined with our multiplexed single-cell data, we constructed a reaction network model of the RAF-MEK-ERK-RSK-S6 and PI3K-AKT-GSK3β-S6 pathways (Mendoza et al., 2011; Manning and Cantley, 2007; Olayioye and Neve, 2000; Wee and Wang, 2017) that included pathway crosstalk at the level of PI3K to RAC (Ebi et al., 2013) and RSK to GSK3β (Manning and Cantley, 2007) (summarized in **Figure 3A**). The reactions were modeled in canonical format to reduce the inclusion of unmeasured reaction components, which is a streamlining step that has worked well in many other contexts (Savageau, 1976; Voit, 2000, 2013). We aggregated all upstream components into a single input with time delay (τ) and an unmeasured modifier variable, pRAF or PI3K, for each pathway. As they were not explicitly modeled, changes that occurred in upstream components (e.g., receptor expression and activation state) or regulation (e.g., altered membrane dynamics) during EMT were represented by changes in parameters, i.e., as magnitudes of pathway inputs.

In contrast to classical model calibration, which uses population averages for parameter estimation, we applied an approach that uses single-cell data to explicitly model single-cell variation and better constrains model parameters. This approach simulates sets of individual cells, where the initial state of each cell is taken from a snapshot measurement, and each combination of the individual cell trajectories represents an empirical multivariate distribution analogous to that derived from multiplexed single-cell measurements (see Methods).

After model fitting, simulations of independently subsampled cells showed strong agreement between model and data in both the marker distributions and parametric statistical features, such as the mean and covariance (**Figures 3B,C**, **Supplementary Figure 3**). Moreover, the shape of the model distributions fit overall very well; departing furthest from measurements in the density of pAKT for mesenchymal cells (**Figure 3B**). This deviation may be due to unmeasured reaction components or modulators that change during EMT, such as membrane-level variances due to cellular morphological switches. Overall, this result confirms that a constant reaction structure is sufficient to represent signaling across EMT.

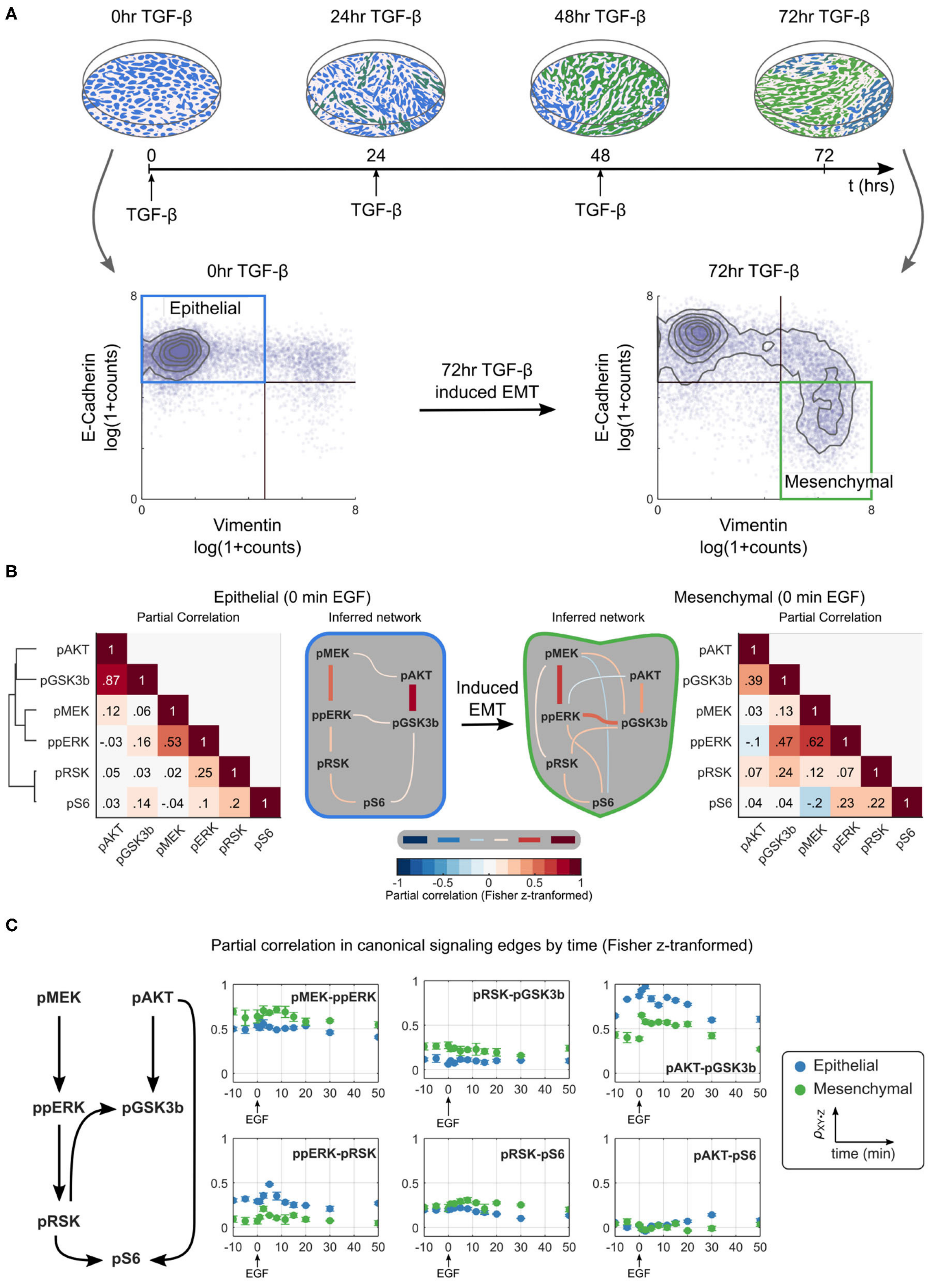


FIGURE 2 | Partial correlation-based network inference suggests that the ERK/AKT signaling network is significantly rewired during EMT. **(A)** Overview of an EMT model (top) including an illustration of the visual phenotype and a molecular definition of epithelial and mesenchymal cells measured by mass cytometry at various time

(Continued)

FIGURE 2 | points (bottom). **(B)** Partial correlation ($\rho_{XY \cdot Z}$) analysis of phosphoproteins is used to infer the different network structures in epithelial and mesenchymal cells sampled across replicates. The threshold for accepting an edge between X and Y is defined as $\rho_{XY \cdot \mathbf{Z}} \geq |0.1|$. All values are Fisher z-transformed. **(C)** Time dependence of signaling strength in canonical signaling edges. Blue and green represent epithelial and mesenchymal phenotypes, respectively. Error bars represent standard deviation of partial correlation across replicates.

2.3. Consolidation of ERK/AKT Signaling in Epithelial and Mesenchymal Cells Requires Only Minimal Adjustments of Mechanistic Model Parameters

Using our unifying reaction network structure, we investigated to what extent the parameters of the dynamic reaction model could be held constant across epithelial and mesenchymal cells. The original point estimates for the model parameter sets differed for the two phenotypes. However, for both phenotypes, there were ranges of parameter values within which the data fit very well. In most cases, these ranges overlapped between the epithelial and mesenchymal cell models. This quality of fit can be visualized in a grid-based sensitivity analysis of how individual parameter changes alter the quality of a model fit (**Figure 4A**). Although these results were encouraging, such univariate sensitivity analyses do not account for the potentially complex effects of changing multiple parameters simultaneously. We therefore systematically searched simultaneously for complete, close-by parameter sets that minimized the differences in parameters between epithelial and mesenchymal cells without decreasing model fitness in either context (**Figures 4B,C**).

As we had hypothesized, based on generic biochemical reasoning, analysis of the mechanistic model revealed that a constant network structure was sufficient to accurately predict signaling alterations during the transition from the epithelial to the mesenchymal state. Amazingly, of the 38 total model parameter values, only four required minor adjustments between epithelial and mesenchymal states: First, I_1, the magnitude of the input to the ERK pathway, is 5.3 in the epithelial state and 1.9 in the mesenchymal state, approaching the pre-EGF steady-state value of 1, at which point there would be no stimulation of the ERK pathway. Second, I_2, the magnitude of the input to the AKT pathway, increases by less than an order of magnitude from 8.1 in the epithelial state to 55 in the mesenchymal state. Third, parameter h_5 decreased from 0.3 in the epithelial state to 0.05 in the mesenchymal state. This parameter is related to the sensitivity of pGSK3β dephosphorylation in response to increases in the concentration of pGSK3β, and accounted for a minute 1% change in model fitness. This change in h_5 may not be surprising given the direct interpretation of the reaction moving from near zero-order (nearly saturated with respect to pGSK3β) to very near zero-order (even closer to saturation), combined with the fact that pGSK3β is generally higher in the mesenchymal state than the epithelial state at across all time points. Fourth, k_{13}, which is related to the rate of S6 phosphorylation, decreases from $10^{-7.8}$ in the epithelial state to $10^{-9.5}$ in the mesenchymal state. This difference could be related to altered expression of unmonitored reaction components or to variations in local concentrations or diffusion rates. Importantly, almost all responses of the system could be explained without evoking natural stochasticity.

To assess the validity of model predicted changes to ERK and AKT pathway inputs, we used a previous molecular study of EMT-dependent signaling across multiple *in vitro* EMT models and more than 70 lung cancer cell lines (Salt et al., 2014). Our model prediction that the ERK pathway input, I_1, decreases across EMT agreed with reported experimental findings that levels of phosphorylated EGFR, as well as total expression of other EGFR family members, were reduced in mesenchymal cell states compared to epithelial cell states. The other relevant prediction of our model, that the AKT pathway input I_2 increased across EMT, also agreed with the experimental findings that PI3KCA, which encodes the catalytic p110α subunit of PI3K upstream of AKT activation, was upregulated and increased AKT pathway signaling across EMT. The consistency of these model predictions with independent experimental observations provides an orthogonal validation of the our model.

3. DISCUSSION

Theoretically, differences in signaling responses across cells should be determined by relative differences in concentrations and states of reaction components, rather than by the idea of substantial reaction network rewiring. We show here that this is indeed the case for signaling in cells before and after the EMT in an *in vitro* model. To demonstrate this consistency, we performed multiplexed single-cell measurements and applied computational modeling to determine how EGF signaling in the ERK and AKT pathways changes with cell phenotype. We generated a large dataset; specifically, we used twice as many measured state variables as had been used previously to create a single-cell mechanistic model of signaling during EMT. Especially valuable for use with the mechanistic modeling approach was the inclusion of markers of total protein to evaluate variation in protein expression.

While our analysis focused on the response of the ERK and AKT pathways to EGF before and after EMT, we note that these two pathways, as well as the SMAD pathway, are involved in TGF-β-induced EMT itself. Furthermore, network-based analyses have shown that signaling via all three pathways is altered during in the same Py2T model of TGF-β-induced EMT used here (Krishnaswamy et al., 2018). Although our results suggest our mechanistic approach can be applied to characterize TGF-β signaling across EMT, our focus was on EGF signaling. The choice of EGF as the acute signaling stimulus over TGF-β was made, first, to avoid potentially confounding input signals due to the chronic TGF-β treatment used induce EMT, and second, due to the relevance of proliferative EGF signaling and potential in the tumor context.

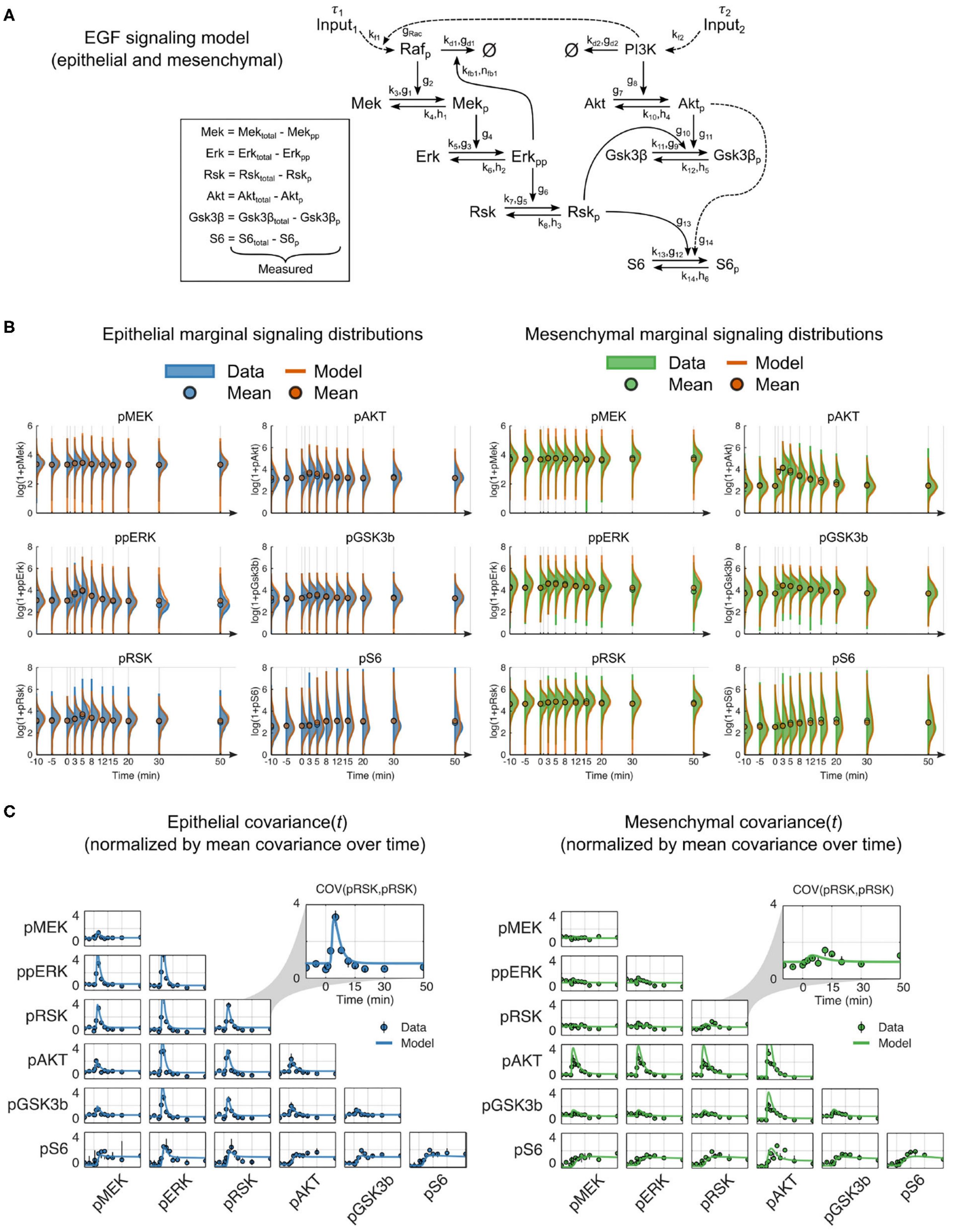


FIGURE 3 | Mechanistic model of EGF signaling in ERK and AKT pathways with parameter fits to epithelial and mesenchymal cells. **(A)** Model reaction structure and measured variables used for each phenotype annotated with kinetic parameters. **(B)** Marginal distributions of (log$_2$ transformed) data (solid) and model simulations

(Continued)

FIGURE 3 | (orange line) for dynamic signaling variables in epithelial (blue solid, left) and mesenchymal (green solid, right) cells. Circles represent means. To enhance visualization, the mean, rather than the marginal distribution, at the 1-min time point is not shown. **(C)** Data (symbols) and model simulations (line) of covariance between signaling variables over time in cells of both phenotypes. Black bars represent the range of covariances across replicates. All values in **(B,C)** were calculated by random subsampling of cells across experimental replicates.

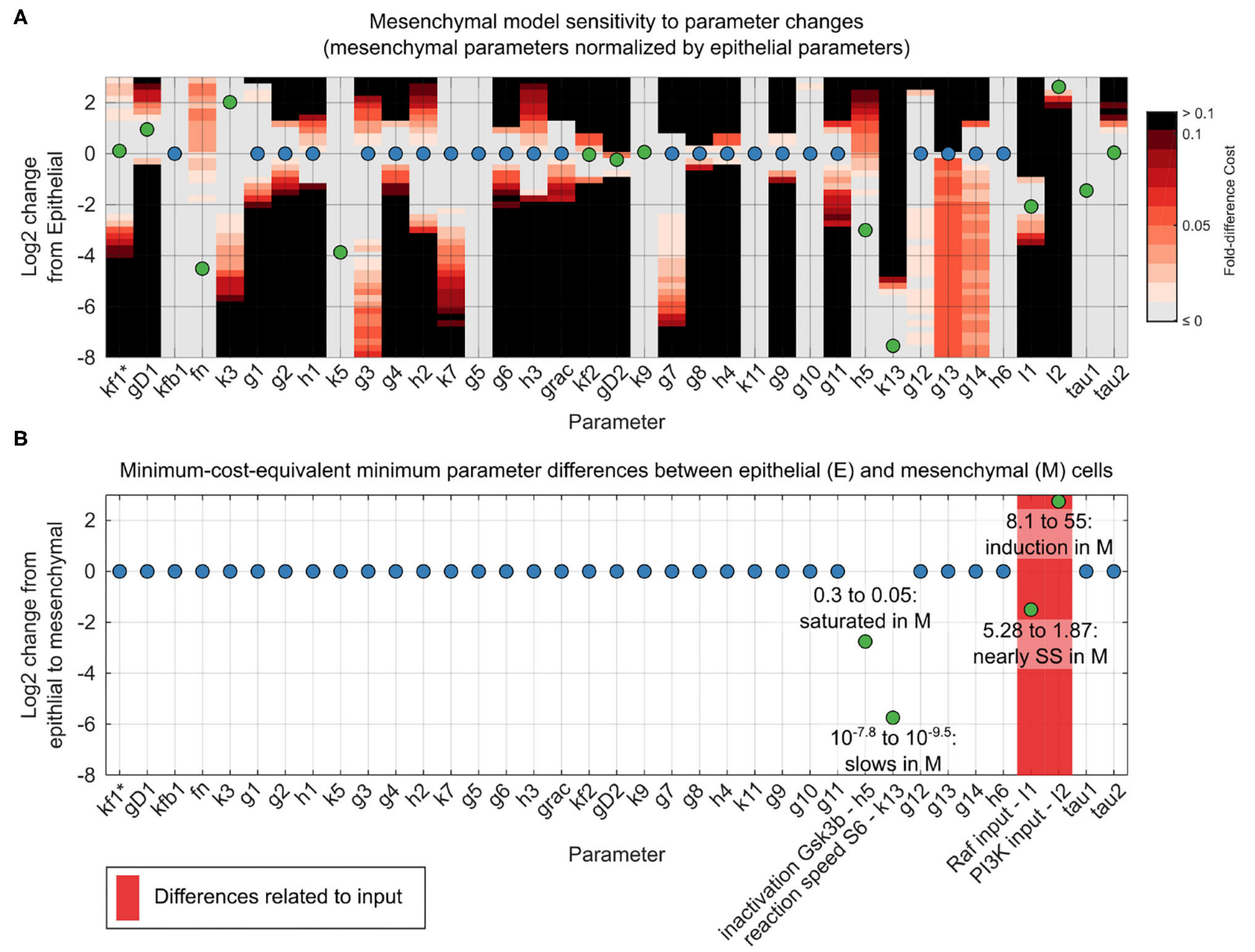


FIGURE 4 | Reconciliation of mesenchymal and epithelial model parameters. **(A)** Sensitivity analysis of mesenchymal model parameter perturbations. Grid represents parameter changes on $\log_2$ scale of epithelial parameters used in **Figure 3**. Circles represent mesenchymal parameters used in **Figure 3** (normalized to epithelial parameters in $\log_2$ scale). If epithelial and mesenchymal parameters are equal, the circles are blue; otherwise they are green. Grid color (color bar) represents percent increase in cost function for the mesenchymal cell model given a corresponding parameter change. **(B)** Minimal parameter difference between epithelial and mesenchymal parameter sets for constant mesenchymal function cost. Parameters associated with model inputs (input magnitude or degradation) are shown in red. Scale and circle colors are as in **(A)**.

We first used a traditional data-driven approach of network inference based on partial correlation, which suggested that the ERK/AKT signaling network had to be rewired in a phenotype-dependent manner during EMT, as was previously reported (Krishnaswamy et al., 2018). In stark contrast, the dynamic, single-cell mechanistic model with constant network structure and near constant kinetic parameter values, calibrated with state-of-the-art single-cell data, predicts that rewiring during EMT is not necessary. Instead, the mechanistic model reconciled the variation in signaling dynamics in epithelial cells vs. mesenchymal cells with only minor EMT-dependent adjustments to model parameters. In summary, a properly calibrated mechanistic model can represent signaling across a contextual change as large as EMT. This mechanistic model suggests that the hypothesis that rewiring of signaling networks does not occur during this phenotypic change is correct; rather cells modulate concentrations of reaction components to alter signaling.

The ability to explain variation in signaling responses from an individual snapshot of cell states is an exciting prospect, but we emphasize that it is necessary to, first, observe dynamics and, second, explicitly measure and model the primary sources of variation in order to obtain accurate predictions. The system components that affect the signaling dynamics will likely include not only levels or functional states of signaling proteins but also features of the microenvironment. For example, if a subpopulation of cells expressed a change of function mutation in a signaling protein, the new molecular species must be added to the model with its own associated kinetic parameters and the expression level must be measured or inferred. Thus, the main limitations of our approach are related to the ability to identify and measure the appropriate variables as well as to the computational cost of fitting mechanistic models to increasingly large systems. These limitations, however, will diminish as multiplexed single-cell measurement technologies and computational infrastructure improve.

Taken together, a single dynamic, mechanistic model with entirely constant reaction structure at the single-cell level can reconcile the EGF signaling dynamics of the ERK/AKT signaling network across an EMT. The only parameter differences we identified associated with the signaling responses of epithelial and mesenchymal phenotypes are slight alterations in four of the 38 total model parameters. Only two of the changes are related directly to kinetic parameters, while the other two are modest variations in the magnitudes of inputs to the pathway. These four EMT-dependent parameters can be interpreted as EMT-dependent changes to reaction components that are not explicitly measured; for example, the increase in I_2 across EMT agrees with the report of an EMT dependent increase in PI3K upstream of AKT (Salt et al., 2014). Thus, we would expect the EMT-dependent variation in I_1 and I_2 to be reduced when including measurements of pEGFR and PI3K. Remarkably, accounting for upstream signaling changes that alter the magnitude of input signal (i.e., I_1 and I_2) to the ERK and AKT pathways was largely sufficient to explain EMT-dependent signaling dynamics, with the inferred EMT-dependent change in the S6 activation rate alone (k_{13}) sufficient to reach at least 99% model fitness for both cell phenotypes.

Our analysis clearly demonstrates that the network structure of murine Py2T breast cancer cells is not rewired during EMT. No mechanisms or processes have to be added and none have to be removed. In fact, it appears that the signaling system is essentially deterministic. Our analysis suggests that the stochasticity inferred from previous studies is by and large a matter of unmeasured variables and a modest degree of intercellular variability. Given the ability of calibrated mechanistic models to analyze, simulate, and explain the dynamics from an initial snapshot of cell states, our data indicate that it is possible to account for cellular context via initialization of an appropriate model with multiplexed single-cell measurements. The proof-of-principle we present here represents an important step toward the construction of tumor-level dynamic models of signaling networks.

4. METHODS

4.1. Cell Culture

Py2T cells were obtained from the laboratory of Gerhard Cristofori, University of Basel, Switzerland; their characterization was previously described (Waldmeier et al., 2012). Cells were tested for mycoplasma contamination upon arrival and regularly during culturing and before being used for experiments. Cells were cultured at 37°C in DMEM (D5671, Sigma Aldrich), supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin, at 5% CO_2. For cell passaging, cells were incubated with TrypLE^{TM} Select 10X (Life Technologies) in PBS in a 1:5 ratio (v/v) for 10 min at 37°C. For each experiment, cells were seeded at the density of 0.3 million cells per plate (100-mm diameter) and allowed to recover for 36 h.

4.2. TGF-β Stimulation

After reaching 60% confluence, cells were either mock-treated or treated with 4 ng/ml human recombinant TGF-β (Cell Signaling Technologies) for 72 h. Medium and TGF-β were replaced every 24 h until 24 h prior to EGF stimulation. For each condition, three biological replicates were cultured, harvested, and analyzed.

4.3. Cell Harvesting and EGF Stimulation

For cell harvest, cells were washed twice with PBS and incubated with TrypLE^{TM} Select 10X (Life Technologies) in PBS at a 1:5 ratio (v/v) for 10 min at 37°C. Following cell detachment, cells were mixed and resuspended in serum-free medium and allowed to recover from detachment for 2 h at 37°C and 5% CO_2 with periodic shaking to avoid cluster formation. After the recovery period, samples were taken to establish baselines. EGF (Peprotech) was then added to a final concentration of 100 ng/ml. Samples were taken at $\{-10, -5, 0, 1, 3, 5, 8, 12, 15, 20, 30, 50\}$ min relative to stimulation ($t = 0$) with EGF (the 0-min sample was not stimulated). At the time of sampling, paraformaldehyde (PFA, from Electron Microscopy Sciences) was added to the cell suspension to a final percentage of 1.6%, and cells were incubated at room temperature for 10 min. Crosslinked cells were washed twice with cell staining medium (CSM, PBS with 0.5% BSA, 0.02% NaN_3). After centrifugation, ice-cold methanol was used to resuspend the cells, followed by a 10-min permeabilization on ice or long-term storage at −80°C.

4.4. Metal-Labeled Antibody Conjugation

The MaxPAR antibody conjugation kit (Fluidigm) was used to generate isotope-labeled antibodies using the manufacturer's standard protocol. After conjugation, the antibody yield was determined based on absorbance at 280 nm. Candor PBS Antibody Stabilization solution was used to dilute antibodies for long-term storage at 4°C.

4.5. Barcoding and Staining Protocol

Formalin-crosslinked and methanol-permeabilized cells were washed three times with CSM and once with PBS. Cells were incubated in PBS containing barcoding reagents (^{102}Pd, ^{104}Pd, ^{105}Pd, ^{106}Pd, ^{108}Pd, ^{110}Pd, ^{113}In, and ^{115}In) at a final concentration of 100 nM for 30 min at room temperature and then washed three times with CSM (Bodenmiller et al., 2012). Barcoded cells were then pooled and stained with the metal-conjugated antibody mix (**Supplementary Table 4**) at room temperature for 1 h. The antibody mix was removed by washing cells three times with CSM and once with PBS. For DNA staining, iridium-containing intercalator (Fluidigm) diluted in PBS with

1.6% PFA was incubated with the cells at 4°C overnight. On the day of the measurement, the intercalator solution was removed, and cells were washed with CSM, PBS, and doubly distilled H_2O. After the last washing step, cells were resuspended in doubly distilled H_2O and filtered through a 70-μm strainer.

4.6. Mass Cytometry Analysis

EQ Four Element Calibration Beads (Fluidigm) were added to cell suspensions in a 1:10 ratio (v/v). Samples were analyzed on a Helios mass cytometer (Fluidigm). The manufacturer's standard operation procedures were used for acquisition at a cell rate of ∼500 cells per second. After the acquisition, all .fcs files from the same barcoded sample were concatenated (Bodenmiller et al., 2012). Data were then normalized, and bead events were removed (Finch et al., 2013) before doublet removal and de-barcoding of cells into their corresponding wells using a doublet-filtering scheme and single-cell deconvolution algorithm (Zunder et al., 2015). Cytobank (http://www.cytobank.org/) was used for additional gating on the DNA channels (^{191}Ir and ^{193}Ir) and ^{139}La/^{141}Pr to remove remaining doublets, debris and contaminating particulates. Data were then exported as .fcs files for subsequent analysis.

4.7. Gating Epithelial and Mesenchymal Cells

Epithelial cells were gated as E-cadherin^{high}/vimentin^{low} in samples without TGF-β treatment. Mesenchymal cells were gated as E-cadherin^{low}/vimentin^{high} in samples treated for 3 days with TGF-β. Gating cutoffs are shown in **Figure 2A**. Treatment with TGF-β for longer than 3 days increases the percentage of Py2T population that undergoes EMT. Samples treated with TGF-β for 3 days also contained cells within the E-cadherin^{high}/vimentin^{low} gate, but these were not considered contextually as epithelial cells.

4.8. Data Normalization and Scaling for Use in Modeling

Experimental measurements were normalized for comparisons across independent experiments and measurements. Cell events with >20 counts for cleaved PARP were removed as dead or apoptotic. Before using variables in modeling, they were linearly scaled to satisfy biological constraints, such as the total units of a protein must be greater than or equal to the units of the phosphorylated form. Cells used in fitting were subsampled across experimental replicates to reduce computational cost. A full description of normalization, scaling, and subsampling of data before use in modeling is presented in the **Supplementary Material**.

4.9. Partial Correlation-Based Network Inference

Given two random variables X and Y and a set of controlling variables $\mathbf{Z} = Z_1, ..., Z_n$, the partial correlation $\rho_{XY\cdot\mathbf{Z}}$ is a measure of the relationship between X and Y when the effects of the $\mathbf{Z} = Z_1, ..., Z_n$ random variables have been accounted for. Mathematically, $\rho_{XY\cdot\mathbf{Z}}$ is the correlation of the residuals e_X and e_Y that result from a linear regression of X and Y with $\mathbf{Z}$, respectively.

To determine the cutoff for partial correlation-based network representations, thresholds were used to define a minimum p-value or correlation coefficient. In order to focus on stronger relationships, we used a threshold based on the partial correlation values. The choice of $|\rho_{XY\cdot\mathbf{Z}}| \geq 0.1$ as the threshold was made as a qualitative boundary between maximizing canonical and minimizing non-canonical signaling relationships in epithelial cells. Most notably, this setting captured the edge between the ERK pathway and pS6 as well as some form of crosstalk between the ERK and AKT pathways. The addition or subtraction of edges based on other thresholds may be readily calculated from the heatmaps provided in **Figure 2**, and in the context of both total and phosphoproteins, in **Supplementary Figure 2**. **Supplementary Figure 2** illustrates the relationship between the partial correlation threshold and network edge number. Heatmap labels were ordered by hierarchical clustering the epithelial population values using single-linkage clustering and Euclidean distance. All partial correlations were calculated using log transformed data values of the form $log(1 + X)$.

4.10. Mechanistic Model of ERK/AKT Pathway Response to EGF

To model EGF signaling dynamics, we used the *Distribution-Free Single-Cell modeling* (DISCO) (Wade et al., 2020) approach described elsewhere in detail. Briefly, cells were assumed to have the same population-level kinetic parameters, which were determined by minimizing the maximum mean discrepancy (MMD) (Gretton et al., 2012), a statistical two-sample test of similarity for n-dimensional distributions, between simulated and experimentally measured distributions. Cell-to-cell variation in unmeasured components was captured in a subset of rate constants that were algebraically determined by a combination of model structure, population-level parameters, and steady-state measurements. To fit individual parameter sets for epithelial and mesenchymal cell populations, we subsampled 500–1,000 cells across the three replicates for each time point. Model equations may be found in the **Supplementary Material**.

4.11. Parameter Optimization

Parameter optimization was performed using a combination of global and local search methods. First, 50,000 initial parameter sets were sampled from a user-input parameter range (see **Supplementary Table 5**). Next, the 200 parameter sets associated with minimum cost were selected. Finally, each parameter set was refined through multiple rounds of optimization using the unconstrained local search algorithm `fminsearch` in the Matlab Optimization Toolbox (The MathWorks, Inc., Natick, MA). The `fminsearch` algorithm begins with a broad search perspective before focusing on a more precise local area. As regular re-initialization of the search improved results, each round of optimization was run for 300 iterations. If, at the end of a round, the cost value had improved by at least 2%, another round was initialized using the parameter set output by the algorithm. Otherwise, the optimization was terminated. After this local search approach, `fminsearch` was initialized a final time with the parameter set corresponding to the lowest cost among the 200 solutions and was run until the model cost stabilized. Point estimates of epithelial and mesenchymal model parameter sets are shown in **Supplementary Table 5**.

4.12. Sensitivity Analysis of Mechanistic Model Parameters

A grid-based sensitivity analysis was performed in the context of the population parameter vector Θ. The cells sampled for model fitting, either epithelial or mesenchymal, and the model structure were held constant. Given the best-fit point estimate of the parameter set for epithelial cells $\Theta^*_E = \theta_{E,1}, ..., \theta_{E,m}$, sensitivities were calculated as the change Δ in cost function F given a change Δ in the jth parameter $\theta_{E,j}$: $\frac{\Delta \mathrm{F}}{\Delta\theta_{E,j}}$ where $\Delta\theta_{E,j}$ was defined using a $\log_2$ fold-change of increments of 0.25 in the range of $[-8, 3]$ when applied to the best parameter $\theta^*_{E,j}$ for epithelial cells. Notably, this definition implies that parameter sensitivities for the epithelial and mesenchymal models were calculated using the same set of values.

4.13. Reconciliation of Mechanistic Model Parameters Across EMT

Reconciliation between the best points estimates for epithelial and mesenchymal population parameter sets Θ^*_E and Θ^*_M, respectively, was performed by finding the minimum difference between all corresponding population parameters θ_E and θ_M that did not increase the cost F associated with either best-fit parameter set:

$$\begin{aligned} \{\hat{\Theta}_{E*,j}, \hat{\Theta}_{M*,j}\} &= \min_j |\Theta^*_{E,j} - \Theta^*_{M,j}| \\ \text{subject to} & \\ F(\hat{\Theta}_{E*}) &\leq F(\Theta^*_E) \\ F(\hat{\Theta}_{M*}) &\leq F(\Theta^*_M) \end{aligned} \tag{1}$$

Point estimates of epithelial and mesenchymal model parameters after reconciliation $(\hat{\Theta}_{E*}, \hat{\Theta}_{M*})$ are shown in **Supplementary Table 5**.

DATA AVAILABILITY STATEMENT

The data set with raw and gated .fcs files and code including model implementation in Matlab and Matlab workspace variables used for modeling are available from Mendeley data: doi: 10.17632/c9h4rx5sf9.

AUTHOR CONTRIBUTIONS

JW, BB, and EV conceptualized the study and wrote the manuscript. JW, X-KL, and NZ designed and performed the experiments. JW conceptualized and performed the analysis. All authors reviewed and edited the final manuscript.

FUNDING

JW was supported by the G. F. Amelio Fellowship awarded by the Georgia Institute of Technology College of Engineering, a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1148903, and a Whitaker International Fellowship awarded by the Institute of International Education. Support also came from [NSF/MCB 1517588 (PI: EV), NSF/MCB (PI: 1411672, PI: D. Downs)], and NIH- 2P30ES019776-05 (PI: C. Marsit). BB's research was funded by an SNSF R'Equip grant, an SNSF Assistant Professorship grant (PP00P3-144874), the European Research Council (ERC) under the European Union's Seventh Framework Program (FP/2007–2013/ERC Grant No. 336921), and NIH (UC4 DK108132).

ACKNOWLEDGMENTS

The authors would like to thank Vito Zanotelli for helpful discussions and Natalie de Souza for fruitful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.579117/full#supplementary-material

REFERENCES

Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., and Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol.* 8, 1195–1203. doi: 10.1038/ncb1497

Altschuler, S. J., and Wu, L. F. (2010). Cellular heterogeneity: do differences make a difference? *Cell* 141, 559–563. doi: 10.1016/j.cell.2010.04.033

Bodenmiller, B., Zunder, E. R., Finck, R., Chen, T. J., Savig, E. S., Bruggner, R. V., et al. (2012). Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol.* 30, 857–866. doi: 10.1038/nbt.2317

Brightman, F. A., and Fell, D. A. (2000). Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. *FEBS Lett.* 482, 169–174. doi: 10.1016/S0014-5793(00)02037-8

Creixell, P., Schoof, E. M., Simpson, C. D., Longden, J., Miller, C. J., Lou, H. J., et al. (2015). Kinome wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* 163, 202–217. doi: 10.1016/j.cell.2015.08.056

Desai, P., Yang, J., Tian, B., Sun, H., Kalita, M., Ju, H., et al. (2015). Mixed-effects model of epithelial-mesenchymal transition reveals rewiring of signaling networks. *Cell. Signal.* 27, 1413–1425. doi: 10.1016/j.cellsig.2015.03.024

Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C., and Healy, J. I. (1997). Differential activation of transcription factors induced by Ca^{2+} response amplitude and duration. *Nature* 386, 855–858. doi: 10.1038/386855a0

Ebi, H., Costa, C., Faber, A. C., Nishtala, M., Kotani, H., Juric, D., et al. (2013). PI3K regulates MEK/ERK signaling in breast cancer via the Rac-GEF, P-Rex1. *Proce. Natl. Acad. Sci. U.S.A.* 110, 21124–21129. doi: 10.1073/pnas.1314124110

Finck, R., Simonds, E. F., Jager, A., Krishnaswamy, S., Sachs, K., Fantl, W., et al. (2013). Normalization of mass cytometry data with bead standards. *Cytometry A* 83, 483–494. doi: 10.1002/cyto.a.22271

Fu, D.-J., Miller, A. D., Southard, T. L., Flesken-Nikitin, A., Ellenson, L. H., and Yu Nikitin, A. (2018). Stem cell pathology. *Annu. Rev. Pathol. Mech. Dis.* 13, 71–92. doi: 10.1146/annurev-pathol-020117-043935

Garmaroudi, F. S., Marchant, D., Si, X., Khalili, A., Bashashati, A., Wong, B. W., et al. (2010). Pairwise network mechanisms in the host signaling response to coxsackievirus B3 infection. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17053–17058. doi: 10.1073/pnas.1006478107

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A Kernel two-sample test. *J. Mach. Learn. Res.* 13, 723–773. Available online at: https://jmlr.csail.mit.edu/papers/v13/gretton12a.html

Halasz, M., Kholodenko, B. N., Kolch, W., and Santra, T. (2016). Integrating network reconstruction with mechanistic modeling to predict cancer therapies. *Sci. Signal.* 9:ra114. doi: 10.1126/scisignal.aae0535

Hill, S. M., Nesser, N. K., Johnson-Camacho, K., Jeffress, M., Johnson, A., Boniface, C., et al. (2017). Context specificity in causal signaling networks revealed by phosphoprotein profiling. *Cell Syst.* 4, 73–83.e10. doi: 10.1016/j.cels.2016.11.013

Kholodenko, B. N. (2006). Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* 7, 165–176. doi: 10.1038/nrm1838

Kolitz, S. E., and Lauffenburger, D. A. (2012). Measurement and modeling of signaling at the single-cell level. *Biochemistry* 51, 7433–7443. doi: 10.1021/bi300846p

Krishnaswamy, S., Zivanovic, N., Sharma, R., Pe'er, D., and Bodenmiller, B. (2018). Learning time varying information ow from single-cell epithelial to mesenchymal transition data. *PLoS ONE* 13:e0203389. doi: 10.1371/journal.pone.0203389

Lee, M. J., Ye, A. S., Gardino, A. K., Heijink, A. M., Sorger, P. K., MacBeath, G., et al. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 149, 780–794. doi: 10.1016/j.cell.2012.03.031

Manning, B. D., and Cantley, L. C. (2007). AKT/PKB signaling: navigating downstream. *Cell* 129, 1261–1274. doi: 10.1016/j.cell.2007.06.009

Mendoza, M. C., Er, E. E., and Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* 36, 320–328. doi: 10.1016/j.tibs.2011.03.006

Olayioye, M., and Neve, R. (2000). The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J.* 19, 3159–3167. doi: 10.1093/emboj/19.13.3159

Petsalaki, E., Helbig, A. O., Gopal, A., Pasculescu, A., Roth, F. P., and Pawson, T. (2015). SELPHI: correlation-based identification of kinase-associated networks from global phospho-proteomics data sets. *Nucleic Acids Res.* 43, W276–W282. doi: 10.1093/nar/gkv459

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P. (2005). Causal protein signaling networks derived from multiparameter single-cell data. *Science* 308, 523–529. doi: 10.1126/science.1105809

Salt, M. B., Bandyopadhyay, S., and McCormick, F. (2014). Epithelial-to-mesenchymal transition rewires the molecular path to PI3K-dependent proliferation. *Cancer Discov.* 4, 186–199. doi: 10.1158/2159-8290.CD-13-0520

Savageau, M. A. (1976). *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Reading, MA: Addison-Wesley Pub. Co., Advanced Book Program.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., et al. (2014). Accurate information transmission through dynamic biochemical signaling networks. *Science* 346, 1370–1373. doi: 10.1126/science.1254933

Voit, E. O. (2000). *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists*. Cambridge: Cambridge University Press.

Voit, E. O. (2013). Biochemical systems theory: a review. *ISRN Biomath.* 2013, 1–53. doi: 10.1155/2013/897658

Wade, J. D., Lun, X. K., Bodenmiller, B., and Voit, E. O. (2020). Multidimensional single-cell modeling of cellular signaling. *bioRxiv [PrePrint]*. doi: 10.1101/2020.11.15.383711

Waldmeier, L., Meyer-Schaller, N., Diepenbruck, M., and Christofori, G. (2012). Py2T murine breast cancer cells, a versatile model of TGFβ-induced EMT *in vitro* and *in vivo*. *PLoS ONE* 7:e48651. doi: 10.1371/journal.pone.0048651

Wee, P., and Wang, Z. (2017). Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers* 9:52. doi: 10.3390/cancers9050052

Will, T., and Helms, V. (2015). PPIXpress: construction of condition-specific protein interaction networks based on transcript expression. *Bioinformatics* 32, 571–578. doi: 10.1093/bioinformatics/btv620

Zunder, E. R., Finck, R., Behbehani, G. K., Amir, E. A. D., Krishnaswamy, S., Gonzalez, V. D., et al. (2015). Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protoc.* 10, 316–333. doi: 10.1038/nprot.2015.020

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 01 December 2020
doi: 10.3389/fphys.2020.594151

Boolean Feedforward Neural Network Modeling of Molecular Regulatory Networks for Cellular State Conversion

Sang-Mok Choo[1], **Laith M. Almomani**[1] **and Kwang-Hyun Cho**[2]*

[1] Department of Mathematics, University of Ulsan, Ulsan, South Korea, [2] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

The molecular regulatory network (MRN) within a cell determines cellular states and transitions between them. Thus, modeling of MRNs is crucial, but this usually requires extensive analysis of time-series measurements, which is extremely difficult to obtain from biological experiments. However, single-cell measurement data such as single-cell RNA-sequencing databases have recently provided a new insight into resolving this problem by ordering thousands of cells in pseudo-time according to their differential gene expressions. Neural network modeling can be employed by using temporal data as learning data. In contrast, Boolean network modeling of MRNs has a growing interest, as it is a parameter-free logical modeling and thereby robust to noisy data while still capturing essential dynamics of biological networks. In this study, we propose a Boolean feedforward neural network (FFN) modeling by combining neural network and Boolean network modeling approach to reconstruct a practical and useful MRN model from large temporal data. Furthermore, analyzing the reconstructed MRN model can enable us to identify control targets for potential cellular state conversion. Here, we show the usefulness of Boolean FFN modeling by demonstrating its applicability through a toy model and biological networks.

Keywords: molecular regulatory network, Boolean network modeling, feedforward neural networks, Boolean feedforward neural network, temporal data, cellular state conversion

Edited by:
Zhike Zi,
Max Planck Institute for Molecular Genetics, Germany

Reviewed by:
Rui Li,
Dalian University of Technology, China
Eberhard Otto Voit,
Georgia Institute of Technology, United States
Jung-Min Yang,
Kyungpook National University, South Korea

****Correspondence:***
Kwang-Hyun Cho
ckh@kaist.ac.kr

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *12 August 2020*
Accepted: *03 November 2020*
Published: *01 December 2020*

Citation:
Choo S-M, Almomani LM and Cho K-H (2020) Boolean Feedforward Neural Network Modeling of Molecular Regulatory Networks for Cellular State Conversion. Front. Physiol. 11:594151. doi: 10.3389/fphys.2020.594151

INTRODUCTION

Cellular behavior is governed by intracellular molecular regulatory networks (MRNs), such as signaling and gene regulatory networks (Schmidt et al., 2005; Kim and Cho, 2006; Sreenath et al., 2008; Kim et al., 2011). Reconstruction and mathematical modeling of such MRNs based on biological experiments have been of great interest in the field of systems biology. Modeling MRNs has been, however, very challenging due to the limited availability of time course measurements from biological experiments. This can now be overcome by recent advancement of technologies in experimental data measurements, and thus, there is a growing interest in developing a new paradigm of modeling MRNs based on large data sets.

Single-cell technologies have emerged in the fields of genomics (Ludwig et al., 2019; Tritschler et al., 2019; Baslan et al., 2020; Yofe et al., 2020), epigenomics (Berkel and Cacan, 2019; Chen et al., 2019; Verma and Kumar, 2019), transcriptomics (Cui et al., 2019; He et al., 2020; Huang et al., 2020), proteomics (Minakshi et al., 2019; Zhu et al., 2019; Labib and Kelley, 2020), and metabolomics (Duncan et al., 2019; Kawai et al., 2019; Kumar et al., 2020). We can now obtain omics information of hundreds to thousands of individual cells from a single experiment. For instance, single-cell RNA sequencing technologies can measure messenger RNA concentration of hundreds to thousands of genes expressed by single cells, and single cell proteomics by mass spectrometry can quantify over 1,000 proteins per single cell at once (Budnik et al., 2018; Lun and Bodenmiller, 2020). Such single-cell data can be used as pseudo-time-series measurements of distinct cellular states that can provide a new opportunity for modeling MRNs.

There have been attempts to develop dynamic models of MRNs based on ordinary differential equations, regression models, and Boolean networks. Boolean models are more appropriate to be employed for modeling MRNs from pseudo-time-series single-cell data since high-throughput single-cell data are more noisy than conventional bulk sequencing data, and Boolean logical network models are relatively robust to noise. Constructing a Boolean network model usually requires two steps: generating pairs of Boolean input and output for each node in the MRN from states of pseudo-time-ordered single cells and then fitting the Boolean state update logic of each node to the data (Hamey et al., 2017). There are, however, a number of challenges in determining the backbone network structure and optimizing the regulatory logic to the measured data sets. To overcome such challenges, we propose an approach combining Boolean network modeling and feedforward neural network (FFN) learning algorithm, which is particularly useful for inferring input–output relationships from large temporal data. For this purpose, we use only temporal data of network nodes and do not need to determine the network structure nor to optimize the regulatory logics. Of note, in our Boolean FFN model, each node of MRN is represented by a single output node of an FFN with all MRN nodes as its input nodes, and then, the state transition dynamics of MRN can be simulated by executing the entire Boolean FFN model.

Considering a cellular state transition process, we can partition the temporal data of such a process into three parts: ordered pairs of initial cellular states, ordered pairs of transitional cellular states, and ordered pairs of final cellular states. These three ordered pairs can then be used for building initial, transitional, and final cellular states of FFNs, which can be referred to as iFFN, tFFN, and fFFN, respectively. Employing the trained iFFN, tFFN, and fFFN, we can generate trajectories starting from initial to terminal cellular states and use such state trajectories as new training data for building a cell fate transition FFN (cFFN) for each node.

The eventual goal of our study is to identify control targets that can induce desired cellular state conversion, and for this purpose, we propose to build cFFN using iFFN, tFFN, and fFFN based on temporal data measurements of network nodes. We demonstrate the effectiveness and possible application of the proposed Boolean FFN modeling of MRNs by applying it to a toy network model as well as real biological networks. In particular, we compare identified control targets for cellular state conversion between the Boolean FFN and its original Boolean network model in order to show the effectiveness of the proposed Boolean FFN modeling of MRNs.

RESULTS

Overview of Constructing cFFN

The overall procedure of constructing a cFFN is summarized in **Figure 1**. We presume that the nodes playing a significant role in the cellular state transition of interest are known, whereas the regulatory relationships among the nodes are unknown (**Figure 1A**). Here, all the nodes are assumed to have binarized values for their expression levels as to consider MRNs represented by Boolean network models. We also assume that marker nodes, which define specific desired or undesired states that are known, can be used as a primary basis for evaluation after identifying control targets for cellular state conversion.

We consider three clusters of Boolean states over the transition from initial to final states through transitional states, resulting in three sets of ordered pairs of initial, transitional, and final cellular states as shown in **Figure 1B**. These will also be referred to as the first, second, and third clusters to emphasize the order of cellular state transition. As we consider a transition process from an initial normal state to a final abnormal state, there is a tendency that the number of desired states decreases from the first to third clusters while the number of undesired states increases, which is referred to as marker tendency. In each cluster, the first state of ordered pair is assumed to be updated to the second state, which is represented by connecting arrows. However, there is no connection information between two clusters, resulting in no trajectory from initial to final states. We call these three consecutive clusters disconnected trajectories.

To construct connected trajectories, we build three FFNs, i.e., iFFN, tFFN, and fFFN, for each node using the corresponding cluster as training data (**Figure 1C**). The marker tendency is used as a constraint for training each FFN.

We consider the first states in the pairs of initial cellular states be the initial input. By applying iFFN, tFFN, and fFFN to each corresponding initial input as iFFN(initial input) and tFFN[iFFN(initial input)], we can construct connected trajectories from initial input to final output states (**Figure 1D**).

Using the set of states on each connected trajectory as new training data, we can construct cFFN for the node (**Figure 1E**). The entire MRN is then composed of cFFNs of the nodes within a network model, which is illustrated by a conceptual diagram in **Figure 1F**.

Toy Network for Illustrating FFNs

Construction of cFFN

We demonstrate an example of building iFFN, tFFN, fFFN, and cFFN using a toy network of six nodes with Boolean update logics to identify control targets in **Figure 2**. The graph in **Figure 2A** only represents collective regulatory relationships between two nodes in the network (without considering the regulatory logics), and node N6 is considered as a unique marker in this case, where a state is the desired state if N6 is active (value 1) or otherwise undesired (value 0) as shown in **Figure 2A**. All possible states

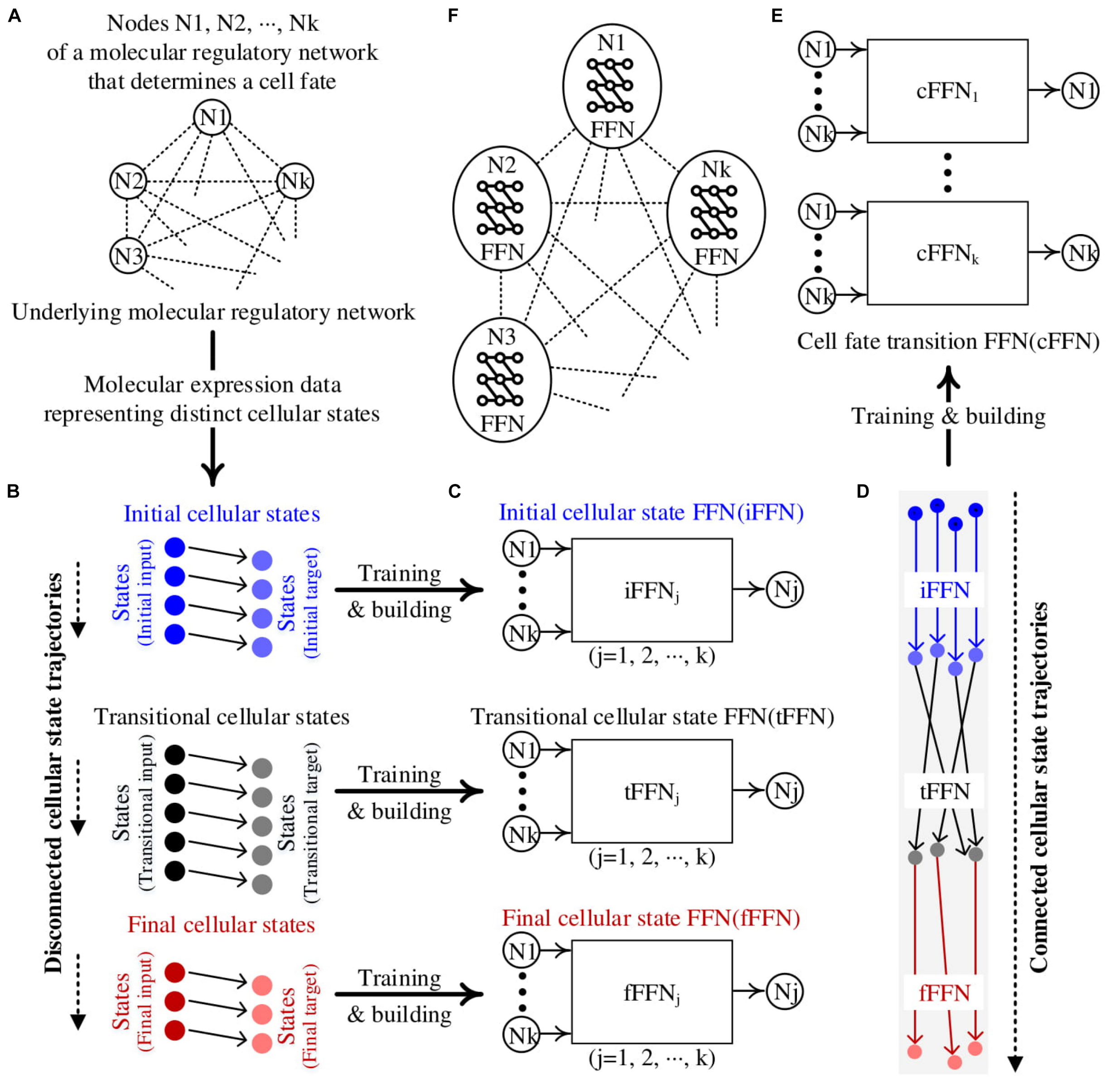


FIGURE 1 | Construction of cell fate transition feedforward neural network (cFFN). **(A)** Nodes represent molecules of the hidden molecular regulatory network (MRN) that can be genes, proteins, or metabolites. **(B)** The measured molecular data representing distinct cellular states are partitioned: ordered pairs of initial, transitional, and final cellular states. Each circle denotes a Boolean state of k nodes for each cellular state. Each right-pointing arrow denotes the direction of updated state. There is no such arrow between two states in different partitions, resulting in disconnected cellular state trajectories. **(C)** The left and right states within the initial cellular states are referred to as initial input and target, respectively, to denote that they are used as training data for an FFN of node Nj. Here, the trained FFN is referred to as an iFFN for the node. Similarly, the left and right states in the transitional and final cellular states are used as training data to build tFFN and fFFN, respectively, for each node. **(D)** The application of iFFN, tFFN, and fFFN to the initial input, iFFN(initial input), and tFFN[iFFN (initial input)], respectively, generates connected trajectories from the initial states. **(E)** States upon each connected trajectory in **(D)** are used as training data for a cFFN. **(F)** A conceptual diagram illustrating our Boolean feedforward neural network modeling. Each circle denotes the constructed Boolean FFN for node Nj.

except one state from the toy network converge to an undesired state, which is called an undesired attractor, and are partitioned into seven sets: D0 denotes a singleton set of the undesired attractor. Dj denotes those states converging to the attractor when they are updated j ($1 \leq j \leq 6$) times.

We use Dj to generate initial, transitional, and final cellular states as shown in **Figure 2B**. Fifteen states randomly chosen from D6, D5, and D4 and their one-time updated states are represented as the first and second states of ordered pairs of initial states, respectively. Fifteen states randomly chosen from D3* and their one-time updated states are represented as the first and second states of ordered pairs of transitional states, respectively. Here, D3* denotes the set of all states in D3 except those states that are updated from the second initial states. Fifteen states

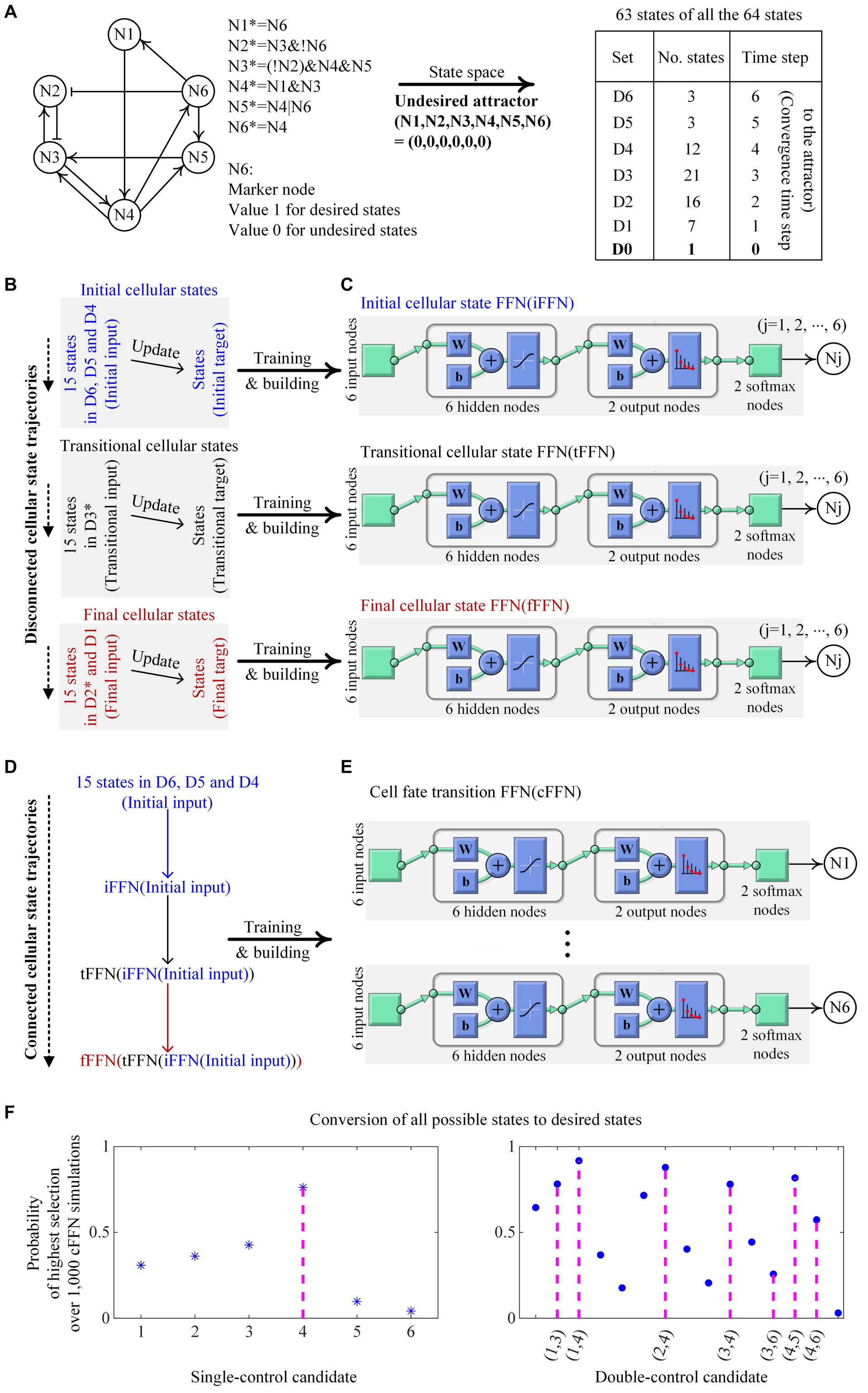


FIGURE 2 | Construction of cFFN with training data generated by a toy network and its application for identifying control targets. **(A)** The toy network consists of nodes Nj ($1 \leq j \leq 6$) with its Boolean update logics. Nj and Nj* denote the Boolean states of Nj at time steps t and t + 1, respectively. Symbols &, |, and ! denote Boolean operators AND, OR, and NOT, respectively. Sharp and blunt arrows represent positive and negative effects, respectively, in the directed graph from Ni to Nj. Node N6 is assumed to be a unique marker, and thus, states of inactive or active N6 correspond to undesired or desired states, respectively. All possible states

(Continued)

FIGURE 2 | Continued
except one state of the network converge to an undesired state, which has all zero values and is designated as a unique undesired attractor. The 63 states are partitioned according to the converging time steps to the undesired attractor where D0 denotes a singleton set of the attractor and Dj denotes a set of states that will become attractors when they are updated j ($1 \leq j \leq 6$) times, referred to as converging time step j to the attractor. **(B)** Initial cellular states denote the states of 15 ordered pairs of which the first states are randomly chosen from D6 to D4 and their updated states become the second states. Transitional and final cellular states are defined similarly, where D3* and D2* denote the sets of all states in D3 and D2 except those states updated from the second initial and transitional states, respectively. Note that the updated states of initial and transitional states are not transitional and final states, respectively. **(C)** By employing the pattern recognition network (PatternNet) and using the training function (train) from Matlab, FFN for each node Nj is trained. The structure of FFN consists of six input nodes, one hidden layer of six nodes, output layer of two nodes, two ordered softmax nodes, and node Nj. Here, softmax nodes have values of the softmax function, and Nj has a value 1 if the value of the first softmax node is greater or equal to that of the second softmax node. The acronyms w and b denote weight and bias, respectively. Each of iFFN, tFFN, and fFFN consists of such six trained FFNs. **(D)** iFFN, tFFN, and fFFN are consecutively applied to the initial input, iFFN(initial input), and tFFN[iFFN(initial input)], and thereby connected cellular state trajectories are produced. **(E)** States upon each connected trajectory are used as training data for an FFN with the structure in **(C)**. The trained FFN is denoted as a cFFN. **(F)** Pinning the value of N4 to 1 is the only way to drive any states to desired ones. Here, N4 is the unique single-control target of value 1. Single-control candidates denote 6 nodes of pinned value 1 in the cFFN. The left panel shows the probability of each single-control candidate of value 1 to be a target of value 1. In this example, the value of Nj is fixed to 1 in the cFFN, and every possible state is updated accordingly. Then, the number of states driven to desired states is counted. Nj gets score 1 if the counted number is in the list of the two highest numbers of the candidates, or 0 otherwise. As a result, repeating this scoring process for each of 1,000 cFFNs gives the probability of Nj in the left panel, where N4 has the highest probability. In the right panel, seven ordered pairs (N1, N3), (N1, N4), (N2, N4), (N3, N4), (N3, N6), (N4, N5), and (N4, N6) of values (1, 1) are the only possible ways to drive any states to desired ones when fixing the values of two nodes to an ordered pair (1, 1). Here, the pairs are double-control targets of values (1, 1) and 15 pairs of two nodes are considered as double-control candidates of values (1, 1) in cFFN. This shows the probability of each double-control candidate of values (1, 1) to be a target of values (1, 1), where the scoring process is the same as that used in identifying single-control target by replacing the two highest numbers with the eight highest numbers.

randomly chosen from D2* and D1 and their updated states are represented as the first and second states of ordered pairs of final states, respectively, where D2* denotes the set of all states in D2 except those states that are updated from the second transitional states (**Supplementary Data 1**).

The first and second states of ordered pairs of initial, transitional, and final states are used for training input and target of iFFN, tFFN, and fFFN, respectively, as shown in **Figure 2C**. The constraint of marker tendency is also considered when training each FFN (see section "Materials and Methods" for details). A sequential application of iFFN, tFFN, and fFFN to the initial input, iFFN(initial input), and tFFN[iFFN(initial input)], produces 15 trajectories as shown in **Figure 2D**. The two consecutive states on each trajectory are used as training input and target for a Boolean FFN, which is cFFN as shown in **Figure 2E**.

Conversion of Undesired States With cFFN

We demonstrate that cFFN can be used in identifying control targets for state conversion of undesired states to desired ones. Pinning the values of single node or two nodes during state update is referred to as single or double controls, respectively. To validate whether the control candidates identified from cFFN can drive the undesired states to desired ones, we compare the control "candidates" to control "targets" found by extensive simulation analysis of the original Boolean network models of MRNs.

Single-control target

To evaluate control candidates, we search for all single-control targets by simulating the Boolean network model of this toy network. For this particular example, when pinning the value of a node to 0 and updating every state according to the regulatory logics of the Boolean network model, there exists a state that cannot be driven to a desired state. This shows that there is no single-control target of value 0 in this case. However, there is a unique single-control target of value 1. Pinning the value of N4 to 1 is the only way to drive all possible states to desired states. This shows that N4 is a unique single-control target of value 1. To examine whether cFFN can be used to identify N4, we consider each node Nj ($1 \leq j \leq 6$) in cFFN as a single-control candidate of value 1.

To identify control candidates as the unique single-control target N4 by using cFFN, we define the probability of each single-control candidate of value 1 to be a single-control target of value 1. Here, the value of Nj is fixed to 1 in a given cFFN, and every possible state is updated using the cFFN. Then, the number of states driven to desired states is counted. After obtaining such counted numbers of all single-control candidates, Nj gets a score 1 if the counted number of Nj is one of the two highest numbers of the candidates, or 0 otherwise. Here, the number 2 is a kind of hyperparameter. We repeat this scoring process for each of 1,000 cFFNs and divide the total score of Nj by 1,000, which is represented as the probability of Nj shown in the left panel of **Figure 2F**. As a result, the single-control target N4 has the highest probability among all of the single-control candidates.

Double-control target

First, we performed a case study for double control by pinning the values of two nodes to (1, 1). We find that, if one of seven pairs, (N1, N3), (N1, N4), (N2, N4), (N3, N4), (N3, N6), (N4, N5), and (N4, N6), has pinned values as (1, 1), any states would eventually converge to desired states. As a result, those seven pairs are identified as double-control targets of values (1, 1). To examine whether cFFN can be used for identifying such double-control targets of values (1, 1), we consider 15 pairs of two nodes as double-control candidates of values (1, 1) and evaluate each of them. To identify control candidates for the double-control targets of values (1, 1) by using cFFN, the probability of each double-control candidate of values (1, 1) to be a target of values (1,1) is defined similarly to that used in the case of single-control candidate. This can be done by replacing single control and the two highest numbers with double control and the eight highest numbers, respectively. We present the probability in the right panel of **Figure 2F**. We find that five of the seven double-control targets of values (1,1) are in the list of five highest probabilities.

We performed the second case study for double control by pinning the values of two nodes to values (0, 1) since there is no double-control target of values (0, 0). If one of five ordered pairs, (N1, N4), (N2, N4), (N3, N4), (N5, N4), and (N6, N4), has values (0, 1), then any states would eventually converge to the desired states. As a result, those five pairs are identified as double-control targets of values (0, 1). To examine whether cFFN can be used for identifying such five double-control targets of values (0,1), we consider 30 ordered pairs of two nodes in cFFN as double-control candidates of values (0, 1) and evaluate each of them. The probability of each double-control candidate of values (0, 1) to be a target of values (0, 1) is defined similarly to that used in the case of double-control candidate of values (1, 1) by replacing the eight highest numbers with the 10 highest numbers. We present the probability in **Supplementary Figure 1**, where all the five double-control targets of values (0, 1) have the five highest probabilities.

Applications of FFN for Identifying Biomolecular Control Targets

To construct cFFN of an MRN and demonstrate its applicability for identifying control targets as in **Figure 3A**, we employ two biomolecular network models. One of the network models is composed of 21 nodes and has a large portion (81.73%) of states converging to an undesired state. In contrast, the other network model is composed of 33 nodes and has a unique undesired state with a very small portion (0.02%) of states converging to an undesired state.

Colitis-Associated Colon Cancer Network

Construction of cFFN

The biomolecular network in **Figure 3B** is a reduced colitis-associated colon cancer network of 21 nodes Nj ($1 \leq j \leq 21$) shown in **Figure 4A**, which is denoted by CACC21 (Lu et al., 2015). Node ID Nj and the state update logics of CACC21 are provided in **Supplementary Data 2**. Markers for desired and undesired states are P53 and Proliferation nodes; states with values (P53, Proliferation) = (1, 0) and (0, 1) are considered as desired and undesired states, respectively. Here, P53 and Proliferation are molecular marker nodes indicating programmed cell death/arrest and uncontrolled cell growth, respectively. In this network, 81.73% of all possible states converge to one of two undesired attractors (Choo et al., 2019 and **Supplementary Data 2**). When we generate 100,000 random states, 81,870 states converge to one of the undesired attractors at time steps from 1 to 11, which are partitioned into 11 sets of Dj ($1 \leq j \leq 11$). D0 denotes a set of the two undesired attractors.

Initial cellular states are 1,000 states randomly chosen from D11 to D9 and their one-time updated states, which are referred to as the first and second states of ordered pairs of initial states, respectively. Here, 800 and 200 states of the first states have values (P53, Proliferation) = (1, 0) and (0, 1), respectively, in **Figure 4B**. Transitional cellular states are 1,000 states randomly chosen from D7 to D6 and their one-time updated states, referred to as the first and second states of ordered pairs of transitional states, respectively. Here, 400 and 600 states of the first states have values (P53, Proliferation) = (1, 0) and (0, 1), respectively. Final cellular states are 1,000 states randomly chosen from D4 to D1 and their updated states, referred to as the first and second states of ordered pairs of final states, respectively. Here, 100 and 900 states of the first states have values (P53, Proliferation) = (1, 0) and (0, 1), respectively.

The first and second states of ordered pairs of initial, transitional, and final states are used as training input and target for iFFN, tFFN, and fFFN, respectively. Restrictions of marker tendency are added for training each FFN (see section "Materials and Methods"). The consecutive application of iFFN, tFFN, and fFFN to the initial input, iFFN(initial input), and tFFN[iFFN(initial input)], respectively, produces 1,000 trajectories that are then used as training data for cFFN.

Single-control target

To validate whether the control candidates identified from cFFN can drive undesired states to desired ones, we compare the control "candidates" to the control "targets" found by extensive simulation analysis of CACC21. We search for all single-control targets by simulating the Boolean network model of CACC21.

There exists no node that can be driven to a desired state when pinning the value of the node to 0 and updating every state according to the regulatory logics of CACC21; there is no single-control target of value 0 in this case. However, there is a unique single-control target of value 1. Pinning the value of N17 (PTEN) to 1 is the only unique single-control target of value 1 that can drive 100 sets of 1,000 random states to desired states. To examine whether cFFN can be used to identify this unique target, we consider each node Nj ($1 \leq j \leq 21$) in cFFN as a single-control candidate of value 1.

To identify control candidates as the unique single-control target N17 by using cFFN, we define the probability of each single-control candidate Nj of value 1 to be a single-control target of value 1. Here, the value of Nj is fixed to 1 in a given cFFN, and 1,000 states of the initial target are updated 100 times using the cFFN. Then, the number of states in the initial target driven to desired states is counted. After obtaining the counted numbers of all single-control candidates of value 1, Nj gets a score 1 if its counted number is one of the five highest numbers of the candidates, or 0 otherwise. We repeat the scoring process for each of 500 cFFNs and divide the total score of Nj by 500, which gives the probability of Nj shown in the upper left panel of **Figure 4C**. Moreover, the initial target used in the scoring process is a hyperparameter. Thus, we replace it with transitional, final, and random states and present the probability of single-control candidates in the upper right, bottom left, and bottom right panels of **Figure 4C**, respectively. As a result, the unique single-control target N17 has also the highest probability.

Double-control target

We find 22 and 30 double-control targets of values (1, 1) and (0, 1), respectively, from extensive simulation analysis of the Boolean network model just as the case of single-control targets. We find that there is a unique double-control target of values (0, 0). Pinning the values of (N8, N13) to (0, 0) is the only way

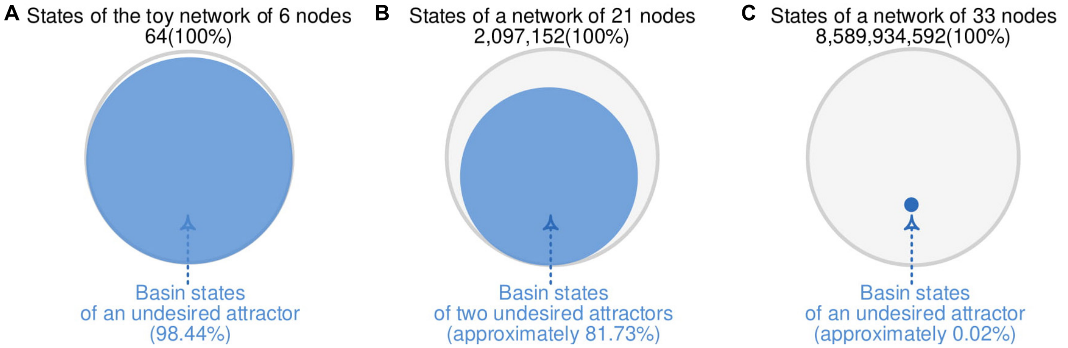


FIGURE 3 | Different portion of states converging to undesired attractors. **(A)** A total of 63 of 64 states from the toy network with six nodes in **Figure 2** converges to an undesired attractor, where the 63 states are referred to as basin states of the undesired attractor with 98.44% basin size. **(B)** A reduced colitis-associated colon cancer network model with 21 nodes in **Figure 4** has 81.73% basin size. **(C)** A mitogen-activated protein kinase network model with 33 nodes in **Figure 5** has 0.02% basin size.

to drive all possible states to desired states, where N8 = IL10 and N13 = MDM2. Hence (N8, N13), is a unique double-control target of values (0, 0). To examine whether cFFN can be used to identify this target, we consider 210 pairs of two nodes as double-control candidates of values (0, 0) and evaluate each of them. To identify control candidates for the double-control targets by using cFFN, we define the probability of each double-control candidate (Nj, Nk) ($1 \leq j < k \leq 21$) of values (0, 0) to be a unique double-control target of values (0, 0). Here, the values of a double-control candidate (Nj, Nk) are fixed to (0, 0) in a given cFFN, and 1,000 states in the final target are updated 100 times by using the cFFN. Then, the number of states driven to desired states is counted. The number of states within the final target driven to desired states is counted for each of 500 cFFNs. After obtaining the counted numbers of all double-control candidates of values (0, 0), (Nj, Nk) gets a score 1 if its counted number is within the top 20% over the total of 210, or 0 otherwise. We repeat this scoring process for each of 500 cFFNs and divide the total score of Nj by 500, which gives the probability of (Nj, Nk) shown in **Supplementary Figure 2**. It shows that the double-control target (N8, N13) of values (0, 0) is within the top 26 out of all 210 probabilities. Moreover, we further test by changing the percentile from top 20% with 30 and 40% and find that the double-control target (N8, N13) of values (0, 0) is within the top 16 and 17, respectively, as shown in **Supplementary Figure 2**.

Mitogen-Activated Protein Kinase Network

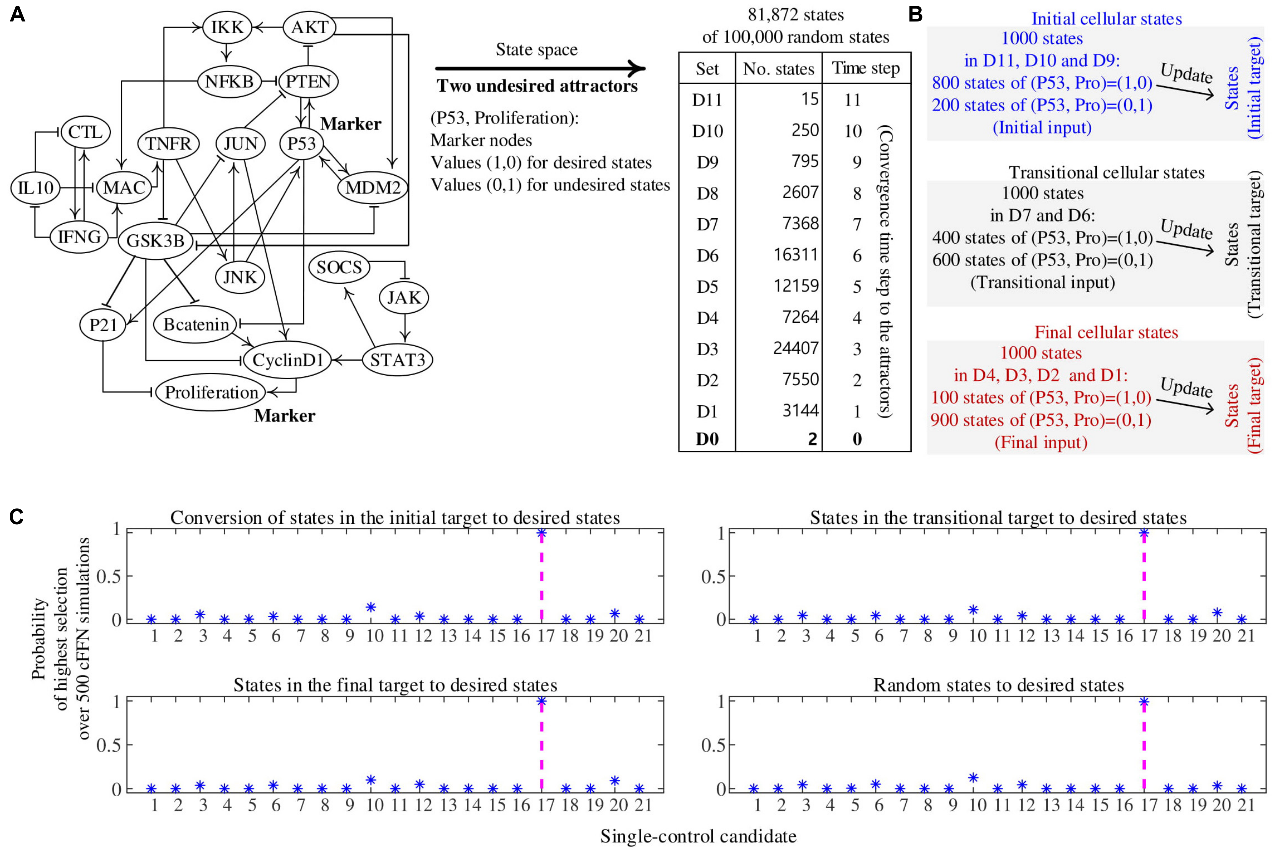
Construction of cFFN

The second biomolecular network in **Figure 3C** is a mitogen-activated protein kinase network composed of 53 nodes as depicted in **Figure 5A** with their update logics in **Supplementary Data 3** (Grieco et al., 2013). The values of epidermal growth factor receptor (EGFR) and all the input nodes are fixed to 1 (dotted diamond in **Figure 5A**) and 0 (dotted rectangles in **Figure 5A**), respectively. As a result, 15 nodes have fixed values (dotted circles in **Figure 5A**). The remaining 33 nodes (solid circles in **Figure 5A**) form a subnetwork, which is called MAPK33. The state update logics of MAKP33 are provided in **Supplementary Data 3**. Nodes of Apoptosis and Proliferation are considered as markers for desired and undesired states; states with (Apoptosis, Proliferation) = (1, 0) and (0, 1) are considered as desired and undesired states, respectively. Here, Apoptosis denotes programmed cell death. About 0.02% out of all possible states converge to an undesired attractor in **Supplementary Data 3 and 4** (Choo et al., 2019).

We generate 1,000 random initial states (first states) converging to the undesired attractor, where 900 and 100 states have active Apoptosis and Proliferation, respectively. The first states are updated one time to the next by using the update logics. Similarly, the second states are updated one time to the third states, which are then also updated to the fourth states. To demonstrate the general applicability of cFFN in identifying control targets, the MAPK33 is used in different ways from the toy network and CACC21 as follows: we introduce noise to the second states by changing the values of six randomly chosen nodes in each of the second states, which are referred to as noisy second states. Similarly, noisy third and fourth states are defined. Therefore, (1st, noisy 2nd), (2nd, noisy 3rd), and (3rd, noisy 4th) are training data for iFFN, tFFN, and fFFN, respectively, and restrictions of marker tendency are added for training each FFN (see section "Materials and Methods" for details). Applying the trained iFFN, tFFN, and fFFN3 to the first states, iFFN(1st states) and tFFN[iFFN1(1st states)], respectively, result in 1,000 connected trajectories from the first states of which are then used to train a cFFN.

Single-control target

To evaluate the control candidates, we search for all single-control targets by simulating the Boolean network model of MAPK33. The first case of single control is pinning the value of one node to 1. There exist four single-control targets of value 1. We find that, if one of four, N12 (GADD45), N20 (MTK1), N24 (P38), and N25 (P53), has the pinned value 1, any state among 2,000 sets of 1,000 random states would eventually converge to a desired state. Here, the states in 1,000 sets are randomly chosen from all the states converging to the undesired attractor. Node ID Nj is provided in **Supplementary Data 3**. As a result, those four nodes are single-control targets of value 1, which are shown



Set	No. states	Time step
D11	15	11
D10	250	10
D9	795	9
D8	2607	8
D7	7368	7
D6	16311	6
D5	12159	5
D4	7264	4
D3	24407	3
D2	7550	2
D1	3144	1
D0	**2**	**0**

FIGURE 4 | Construction of cell fate transition feedforward neural network (cFFN) with training data generated by a biomolecular network and its application to identifying single-control targets. **(A)** A reduced colitis-associated colon cancer network of 21 nodes (CACC21). Nodes P53 and Proliferation are considered as markers for desired and undesired states, respectively. Its update logics and two undesired attractors are shown in **Supplementary Data 2**. We generate 100,000 random states, of which 81,872 states converge to one of the undesired attractors and are partitioned into Dj ($1 \leq j \leq 11$). Dj denotes a set of states that will transit into one of the attractors when they are updated j times, which is referred to as converging time step j to the attractors. D0 denotes a set of the two undesired attractors. **(B)** The 1,000 states in the initial cellular states denote 1,000 randomly generated states from D11, D10, and D9. Here, 800 states have (P53, Proliferation) = (1,0) and 200 states have (P53, Proliferation) = (1,0), which are referred to as initial input for training iFFN. The states are updated one time, where the updated states are referred to as initial target for training iFFN. The 1,000 states in the transitional cellular states denote 1,000 randomly generated states from D7 and D6. Here, 400 states have (P53, Proliferation) = (1,0) and 600 states have (P53, Proliferation) = (1,0), which are referred to as transitional input for training tFFN. The transitional target denotes one-time updated states in the transitional input. The 1,000 states in the final cellular states denote 1,000 randomly generated states from D4, D3, and D2. Here, 100 states have (P53, Proliferation) = (1,0) and 900 states have (P53, Proliferation) = (1,0), which are referred to as final input for training fFFN. Similarly, final target is defined by using final input. **(C)** Four subplots show the probability of each single-control candidate of value 1 to be a single-control target, where 500 cFFNs and the five highest numbers are used in the scoring process. The states in the initial, transitional, and final targets and random states are used in the scoring process. As a result, the unique single-control target N17 of value 1 has the highest probability.

as red nodes in **Figure 5A**. To examine whether cFFN can be used to identify these single-control targets of pinning value 1, we consider each node Nj ($1 \leq j \leq 33$) in cFFN as a single-control candidate of value 1.

To identify control candidates as a single-control target by using cFFN, we define the probability of single-control candidate Nj of value 1 to be a single-control target of value 1. Here, the value of Nj is fixed to 1 in a given cFFN, and the 1,000 noisy fourth states are updated 100 times using the cFFN. Then, the number of noisy fourth states driven to desired states is counted. After obtaining such counted numbers of all single-control candidates, Nj gets a score 1 if the counted number of Nj is 1 of the 10 highest numbers among the candidates, or 0 otherwise. We repeat the scoring process for each of 100 cFFNs and divide the total score of Nj by 100, which gives the probability of Nj shown in the upper panel of **Figure 5B**. As a result, the four single-control targets of value 1 are ranked as the third, fifth, ninth, and second in descending order of probability.

The middle panel of **Figure 5B** shows the probability of each single-control candidates to be a single-control target of value 1 without introducing the restriction of marker tendency. In addition, 1,000 states at converging time step 1 to the undesired attractor are used instead of the noisy fourth states driven to desired states in the upper panel of **Figure 5B**. As a result, the four single-control targets are ranked as the third, third, fifth, and second in descending order of probability. Finally, the bottom panel in **Figure 5B** shows the probability of each single-control candidates to be a single-control target of value 1. This

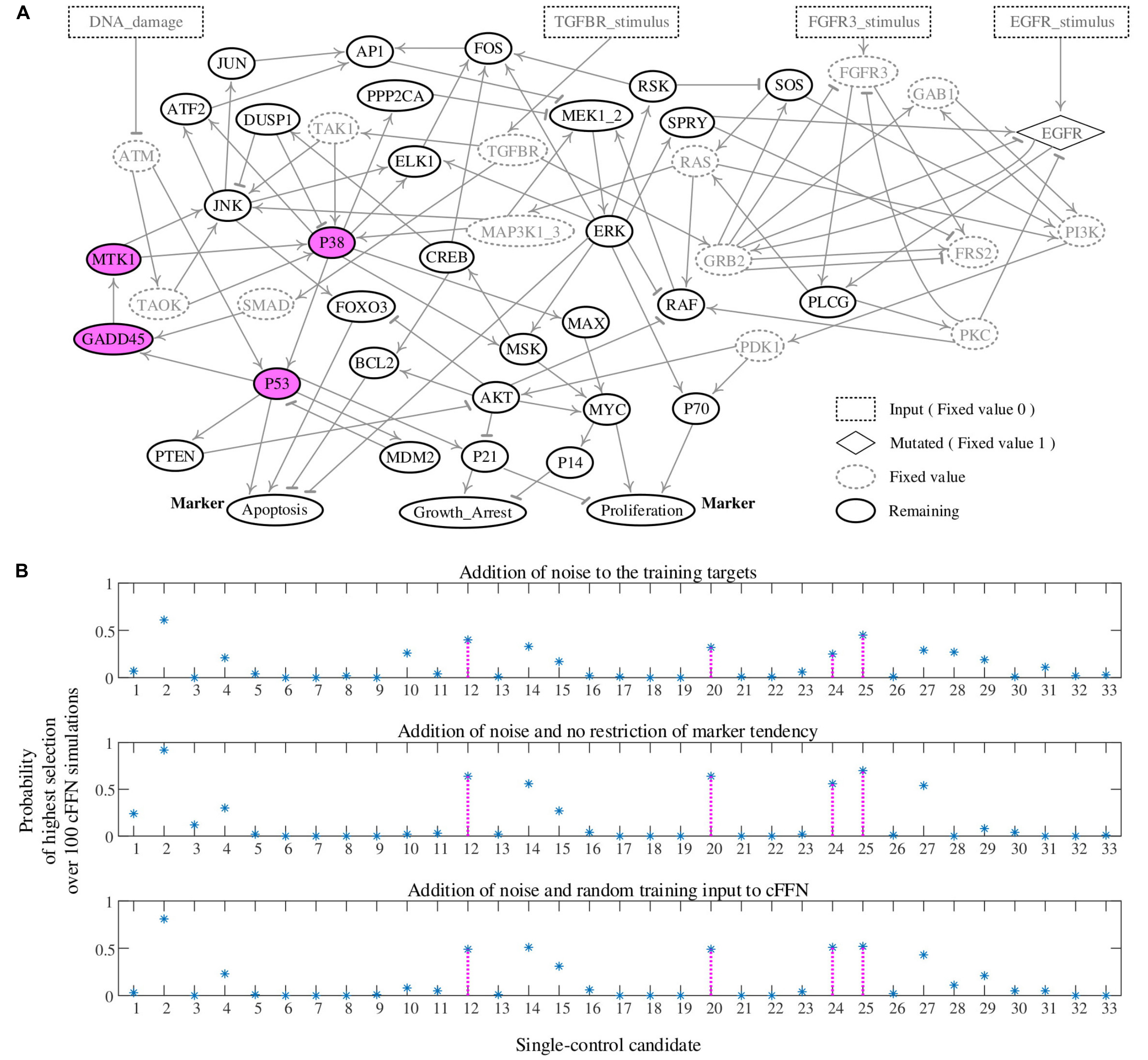


FIGURE 5 | Construction of cell fate transition feedforward neural network (cFFN) with noisy training data generated by a biomolecular network of 33 nodes and application of cFFN for identifying single-control targets of value 1. Node ID Nj ($1 \leq j \leq 33$) is provided in **Supplementary Data 3**. **(A)** A mitogen-activated protein kinase network of 53 nodes. Its update logics are provided in **Supplementary Data 3**. Rectangle and diamond nodes have fixed values of 0 and 1 in the update logics, respectively, which lead to fixed values of the dotted circles. All nodes except the nodes having fixed values are 33 nodes, which are represented by solid circle nodes. The network of 33 nodes is called MAPK33. Update logics of these nodes are provided in **Supplementary Data 3**. If one of four, N12, N20, N24, and N25, has the pinned value 1, any state within 2,000 sets of 1,000 random states would eventually converge to a desired state, where states in 1,000 sets are randomly chosen from all the states converging to the undesired attractor. The four red nodes denote the four single-control targets of value 1. **(B)** In the upper panel, the probability of single-control candidate Nj to be a single-control target of value 1 is calculated using the noisy training targets for iFFN, tFFN, and fFFN. Here, restrictions of marker tendency and the number of noisy fourth states driven to desired states are used. The four single-control targets of value 1 are ranked as the third, fifth, ninth, and second in descending order of probability. In the middle, the probability is calculated using the noisy training targets for iFFN, tFFN, and fFFN. Moreover, no restrictions of marker tendency and 1,000 states at time step 1 to the undesired attractor instead of the noisy fourth states are used. As a result, the four single-control targets are ranked as the third, third, fifth, and second in descending order of probability. In the bottom panel, the probability is calculated using noisy training targets for iFFN, tFFN, and fFFN, and random states converging to the undesired attractor and the noisy first states instead of the first states and the noisy fourth states, respectively. Finally, the four single-control targets are ranked as the fifth, fifth, third, and second in descending order of probability.

can be done by using 1,000 random states converging to the undesired attractor and the noisy second states, instead of the first states and the noisy fourth states in the upper panel of **Figure 5B**, respectively. As a result, the four single-control targets are ranked as the fifth, fifth, third, and second in descending order of probability.

The second case of single control is pinning the value of one node to value 0. We find that if one of two nodes, N6 (CREB) and N7 (DUSP1), has value 0, then any state within 100 sets of 1,000 random states converging to the unique undesired attractor would eventually converge to desired states. Hence, the two nodes are single-control targets of value 0. We define the probability of single-control candidate Nj to be a single-control target of value 0 as in the upper panel of **Figure 5B**. As a result, the two single-control targets of value 0 have the first two highest probabilities as shown in **Supplementary Figure 3**.

DISCUSSION

Owing to the recent development of high throughput single cell measurement technologies, various omics data are now becoming more available that can be used for quantifying gene or protein expressions of hundreds to thousands of cells at a time. Those data can be ordered according to pseudo-time axis, and then, we can use them to investigate dynamic processes in cellular state transitions such as differentiation and tumorigenesis. One most important application of such data is developing a mathematical model of the MRN within a cell since it determines cellular dynamic behaviors. Boolean network models have been actively studied, as they are parameter-free logical dynamic models that can still represent many essential dynamics of MRNs and are also robust to noise contained in the data used for logic fitting. All previous studies on developing Boolean network models have focused on inferring the backbone network structures and optimizing the regulatory logical rules among the nodes of MRNs. In this study, we proposed a totally different approach by representing each node (i.e., molecule) of MRNs by a single output node of an FFN and then fitted the whole MRN composed of as many FFNs as the number of nodes to the measured pseudo-time course data such that the resulting Boolean FFN can reproduce all the predicted molecular expression levels of nodes for any initial state values. In this approach, we do not need to determine the regulatory network structure in advance, as it is obtained as a result of learning. To use our method, we only need to know (or determine) *a priori* the nodes that constitute the regulatory network. Then, we can apply our method based on the temporal measurement data of the network nodes. It is also remarkable that the proposed Boolean FFN modeling is quite robust to noise in the training data.

To show validity and applicability of the proposed Boolean FFN modeling, we considered three different network examples and further applied our method to identify control targets that can induce cellular state conversion to desired ones. We found that our method can accurately identify all those control targets that are revealed by extensive simulation analysis of the original dynamic network models. For the toy example network and CACC21 network, three clusters of states that are sequentially ordered upon the pseudo-time course trajectory leading to undesired attractors are generated by dividing the state transition trajectory into initial, middle, and final clusters of states. The first cluster is a set of states at early time steps, where ordered pairs of the states and their updated states are used as training data for iFFN. Similarly, the second and third clusters are defined and used as training datasets for tFFN and fFFN, respectively. Finally, using the first cluster and three FFNs, connected trajectories are constructed and the states on which are used as training data for cFFN. We used an ensemble of cFFNs to identify control targets for cellular state conversion, which shows general applicability of the proposed Boolean FFN modeling to biological network control for state conversion. Identifying control targets is important for cell fate control toward a desired cellular state. For instance, we can consider a state conversion from a malignant cancerous state to a benign normal state, which is called cancer reversion (Cho et al., 2016, 2017; Choi et al., 2020; Lee et al., 2020). We also showed that our method is robust to noisy data through the example of MAPK33 network.

Three Boolean FFNs (iFFN, tFFN, and fFFN) are used only to generate training data for building cFFN, but they can also be used to identify control targets for cellular state conversion among the initial, transitional, and final cellular states. Nevertheless, the key aspect of our proposed framework does not lie in the concept of iFFN, tFFN, and fFFN but in combining neural network modeling and Boolean network modeling approaches. In other words, we can use temporal data as training data for directly building a Boolean FFN without building iFFN, tFFN, and fFFN. Note that we investigated the attractor of a Boolean network only to generate temporal data, so our framework can be used without searching for attractors if temporal data of a cellular state transition process are given. To demonstrate the applicability of our framework without building such three Boolean FFNs and searching for an attractor, we employed the actually measured pseudo-time course single-cell data over the progression from hematopoietic stem cells toward lymphoid-primed multipotent progenitors (Hamey et al., 2017) and directly built a Boolean FFN using these pseudo-time data. From the FFN, we could identify an optimal single-control candidate as shown in **Supplementary Figure 4**. It remains as a future study for experimentally validating this result. Our future study also includes applying the proposed framework to identifying control candidates for cancer reversion together with its experimental validation.

We note that the proposed method might fail to identify optimal control targets if the training data are randomly chosen from a set of small portion of states having a property converging to an undesired attractor. We also note that there are many hidden hyperparameters to be determined in our proposed modeling framework since we employed a machine learning algorithm, FFN. For instance, we used only one hidden layer for the structure of FFN, and the number of hidden nodes was simply set to the number of molecules in the MRN. Although the structure is very simple, it worked well for the temporal data obtained from both the toy model and biological networks. However, different structures and hyperparameter values might be chosen for temporal data from other biological networks. The proposed Boolean FFN modeling framework of MRNs is universal, and thus, it can be applied to various types of molecular data obtained across state transitions.

MATERIALS AND METHODS

Building Feedforward Neural Networks With Matlab

Let us consider an MRN represented by a Boolean network model. Here, we propose a new approach for modeling each node $x_i\ (1 \le i \le k)$ of the Boolean network model using FFN. For this, we assume that three sets of ordered pairs of initial, transitional, and final states are measured over a dynamic process of state transition, which are denoted by PI, PT, and PF, respectively. These are defined as follows:

$$PI = \left\{ \left(s^I_{in,n}, s^I_{tar,n}\right) \middle|\ 1 \le n \le I_k,\ \left\{s^I_{in,n}, s^I_{tar,n}\right\} \subset \{0,1\}^k \right\},$$
$$PT = \left\{ \left(s^T_{in,n}, s^T_{tar,n}\right) \middle|\ 1 \le n \le T_k,\ \left\{s^T_{in,n}, s^T_{tar,n}\right\} \subset \{0,1\}^k \right\},$$
$$PF = \left\{ \left(s^F_{in,n}, s^F_{tar,n}\right) \middle|\ 1 \le n \le F_k,\ \left\{s^F_{in,n}, s^F_{tar,n}\right\} \subset \{0,1\}^k \right\}.$$

where PI has I_k ordered pairs $\left(s^I_{in,n}, s^I_{tar,n}\right)$ for $1 \le n \le I_k$ such that $s^I_{in,n}$ and $s^I_{tar,n}$ are Boolean states of nodes $x_i\ (1 \le i \le k)$. The symbols n and I in $\left(s^I_{in,n}, s^I_{tar,n}\right)$ denoted that $\left(s^I_{in,n}, s^I_{tar,n}\right)$ is an nth pair of PI. The symbols in and tar in $\left(s^I_{in,n}, s^I_{tar,n}\right)$ denoted that $s^I_{in,n}$ and $s^I_{tar,n}$ are elements of input and target for training an FFN, respectively. The symbols in defining the terms, PT and PF, are similar to that ofPI.

Construction of an FFN With Training Data *PI* (iFFN)

Using the Matlab function "patternnet," we construct the structure of FFN for a node x_i with input, one hidden layer, one output layer of two nodes, two softmax nodes and node x_i. The value of x_i is 1 if the value of the first softmax node out of two is greater than or equal to that of the second node. The FFN for node x_i is trained with input $s^I_{in} = \left(s^I_{in,1}, \cdots, s^I_{in,I_k}\right)$ and target $s^I_{tar,x_i} = \left(s^I_{tar,1,x_i}, \cdots, s^I_{tar,I_k,x_i}\right)$ by using the Matlab function "train:"

$s^I_{tar,\ell,x_i} = \left(s^I_{tar,\ell,x_i,1},\ s^I_{tar,\ell,x_i,2}\right)$ such that

$$s^I_{tar,\ell,x_i,1} = \begin{cases} 1 \text{ if } x_i \text{ has value 1 in } s^I_{tar,\ell}, \\ 0 \text{ otherwise,} \end{cases}$$
$$s^I_{tar,\ell,x_i,2} = 1 - s^I_{tar,\ell,x_i,1}.$$

The sizes of input and target are $k \times I_k$ and $2 \times I_k$, respectively. For training, we use the classification threshold of 0.6 and add restrictions that the FFN can satisfy the marker tendency, which is described in detail in section "Marker Tendency." The trained FFN for node x_i is called "$FFN^1_{x_i}$." Then, we can use a vector-valued function notation as follows:

$iFFN = \left(FFN^1_{x_1}, \cdots, FFN^1_{x_k}\right)$ and simply $F^1 = iFFN$.

Then, the output of F^1 can be written as follows:

$$F^1\left(s^I_{in}\right) = \left\{ \left(FFN^1_{x_1}(s), \cdots, FFN^1_{x_k}(s)\right) \middle|\ s \in s^I_{in} \right\},$$

where $FFN^1_{x_i}(s)$ denotes the value of x_i obtained by substituting a state $s \in s^I_{in}$ into $FFN_{x_i}{}^1$.

Construction of an FFN With Training Data *PT* (tFFN)

Using the Matlab function "patternnet," we construct the structure of FFN for a node x_i with input, one hidden layer, one output layer of two nodes, two softmax nodes, and node x_i. The value of x_i is 1 if the value of the first softmax node out of two is greater than or equal to that of the second node. The FFN for node x_i is trained with input $s^T_{in} = \left(s^T_{in,1}, \cdots, s^T_{in,T_k}\right)$ and target $s^T_{tar,x_i} = \left(s^T_{tar,1,x_i}, \cdots, s^T_{tar,T_k,x_i}\right)$ by using the Matlab function "train:"

$s^T_{tar,\ell,x_i} = \left(s^T_{tar,\ell,x_i,1},\ s^T_{tar,\ell,x_i,2}\right)$ such that

$$s^T_{tar,\ell,x_i,1} = \begin{cases} 1 \text{ if } x_i \text{ has value 1 in } s^T_{tar,\ell}, \\ 0 \text{ otherwise,} \end{cases}$$
$$s^T_{tar,\ell,x_i,2} = 1 - s^T_{tar,\ell,x_i,1}.$$

The sizes of input and target are $k \times T_k$ and $2 \times T_k$, respectively. For training, we use the classification threshold of 0.6 and add restrictions that the FFN can satisfy the marker tendency. The trained FFN is called "$FFN^2_{x_i}$." Then, we can use a vector-valued function natation as follows:

$tFFN = \left(FFN^2_{x_1}, \cdots, FFN^2_{x_k}\right)$ and simply $F^2 = tFFN$.

Construction of an FFN With Training Data *PF* (fFFN)

Using the Matlab function "patternnet," we construct the structure of FFN for a node x_i with input, one hidden layer, one output layer of two nodes, two softmax nodes, and node x_i. The value of x_i is 1 if the value of the first softmax node out of two is greater than or equal to that of the second node. The FFN for node x_i is trained with input $s^F_{in} = \left(s^F_{in,1}, \cdots, s^F_{in,F_k}\right)$ and target $s^F_{tar,x_i} = \left(s^F_{tar,1,x_i}, \cdots, s^F_{tar,F_k,x_i}\right)$ by using the Matlab function "train:"

$s^F_{tar,\ell,x_i} = \left(s^F_{tar,\ell,x_i,1},\ s^F_{tar,\ell,x_i,2}\right)$ such that

$$s^F_{tar,\ell,x_i,1} = \begin{cases} 1 \text{ if } x_i \text{ has value 1 in } s^F_{tar,\ell}, \\ 0 \text{ otherwise,} \end{cases}$$
$$s^F_{tar,\ell,x_i,2} = 1 - s^F_{tar,\ell,x_i,1}.$$

The sizes of input and target are $k \times F_k$ and $2 \times F_k$, respectively. For training, we use the classification threshold of 0.6 and add restrictions that the FFN can satisfy the marker tendency. The trained FFN is called "$FFN^3_{x_i}$." Then, we can use a vector-valued function notation as follows:

$fFFN = \left(FFN^3_{x_1}, \cdots, FFN^3_{x_k}\right)$ and simply $F^3 = fFFN$.

Construction of a Cell Fate Transition FFN

We call the following set as "connected trajectories:"

$$\{(s_1, s_2, s_3, s_4) \mid (s_1, s_2, s_3, s_4) \in s^I_{in} \times F^1\left(s^I_{in}\right) \times F^2\left(F^1\left(s^I_{in}\right)\right) \times$$
$$F^3\left(F^2\left(F^1\left(s^I_{in}\right)\right)\right).\},$$

where the symbol $\in$ represents that s_1, s_2, s_3 and s_4 are one of states$s^I_{in,n}$, $FFN^1_{x_i}\left(s^I_{in,n}\right)$, $FFN^2_{x_j}\left(FFN^1_{x_i}\left(s^I_{in,n}\right)\right)$, and $FFN^3_{x_\ell}\left(FFN^2_{x_j}\left(FFN^1_{x_i}\left(s^I_{in,n}\right)\right)\right)$ for $1 \le n \le I_k$ and $1 \le x_i, x_j, x_\ell \le k$, respectively. An FFN for node x_i is trained with input $s^{cFFN}_{in} = \left(s^I_{in},\ F^1\left(s^I_{in}\right),\ F^2\left(F^1\left(s^I_{in}\right)\right)\right)$ and target

$$s^{cFFN}_{tar,x_i} = \left(F^1\left(s^I_{in}\right)_{x_i},\ F^2\left(F^1\left(s^I_{in}\right)\right)_{x_i},\ F^3\left(F^2\left(F^1\left(s^I_{in}\right)\right)\right)_{x_i}\right)$$

by using the Matlab function "train." Here,

$$F^1\left(s^I_{in}\right)_{x_i} = \left(F^1\left(s^I_{in,1}\right)_{x_i}, \cdots, F^1\left(s^I_{in,I_k}\right)_{x_i}\right) \text{ such that}$$

$$F^1\left(s^I_{in,n}\right)_{x_i} = \left(F^1\left(s^I_{in,n}\right)_{x_i,1}, F^1\left(s^I_{in,n}\right)_{x_i,2}\right) \text{ and}$$

$$F^1\left(s^I_{in,n}\right)_{x_i,1} = \begin{cases} 1 \text{ if } x_i \text{ has value 1 in } F^1\left(s^I_{in,n}\right), \\ 0 \text{ otherwise,} \end{cases}$$

$$F^1\left(s^I_{in,n}\right)_{x_i,2} = 1 - F^1\left(s^I_{in,n}\right)_{x_i,1}.$$

The remaining symbols in $F^2\left(F^1\left(s^I_{in}\right)\right)_{x_i}$ and $F^3\left(F^2\left(F^1\left(s^I_{in}\right)\right)\right)_{x_i}$ are similarly defined as those in $F^1\left(s^I_{in}\right)_{x_i}$. For training, we use the classification threshold of 0.6 and add restrictions that the FFN satisfies the marker tendency, which is described in section "Marker Tendency." The trained FFN is called "$cFFN_{x_i}$." Then, we can use a vector-valued function notation as follows:

$$cFFN = \left(cFFN_{x_1}, \cdots, cFFN_{x_k}\right),$$

which is called a "cellular state transitional FFN (cFFN)."

Marker Tendency

We assume that there are marker nodes in the network that can define a state as desired or undesired state. In addition, there is a tendency that the number of desired states in training data decreases from iFFN to tFFN and then to fFFN. However, the number of undesired states in training data increases, which is referred to as marker tendency. We impose restrictions on marker tendency for training iFFN, tFFN, fFFN, and cFFN. We use symbol $\#FFN^\xi_{x_i}\left(s^\tau_{in}\right)$ to denote the number of states with active x_i in the output $FFN^\xi_{x_i}\left(s^\tau_{in}\ \right)$.

Toy Network

Node x_6 is a unique marker; a state with active or inactive x_6 is considered as a desired or undesired state, respectively. Let $\#s^I_{in}$ and $\#s^I_{tar}$ denote the number of desired states in s^I_{in} and $s^I_{tar} = \left(s^I_{tar,1}, \cdots, s^I_{tar,I_k}\right)$, respectively. When training $FFN^1_{x_6}$, we use a lower bound and an upper bound

$lower^I_{x_6} = \frac{1}{2}\min\left\{\#s^I_{in}, \#s^I_{tar}\right\}$ and $upper^I_{x_6} = 2lower^I_{x_6}$ to add the restriction of marker tendency

$$lower^I_{x_6} \le \#FFN^1_{x_6}\left(s^I_{in}\right) \le upper^I_{x_6}.$$

Replacing $\left(\#s^I_{in}, \#s^I_{tar}\right)$ with $\left(\#s^T_{in}, \#s^T_{tar}\right)$, we add the restriction of marker tendency when training $FFN^2_{x_6}$:

$$lower^T_{x_6} \le \#FFN^2_{x_6}\left(s^T_{in}\right) \le upper^T_{x_6}.$$

Replacing $\left(\#s^I_{in}, \#s^I_{tar}\right)$ with $\left(\#s^F_{in}, \#s^F_{tar}\right)$ and $\left(\#s^{cFFN}_{in}, \#s^{cFFN}_{tar}\right)$, we define $upper^F_{x_6} = \frac{1}{2}\min\left\{\#s^F_{in}, \#s^F_{tar}\right\}$ and $upper^{cFFN}_{x_6} = \frac{1}{2}\min\left\{\#s^{cFFN}_{in}, \#s^{cFFN}_{tar}\right\}$.

We add the restrictions of marker tendency when training $FFN^3_{x_6}$:

$$\#FFN^3_{x_6}\left(s^C_{in}\right) \le upper^F_{x_6},\ \#cFFN_{x_6}\left(s^I_{in}\right) \le upper^{cFFN}_{x_6}.$$

CACC21

Nodes x_{16} and x_{21} are the markers for desired and undesired state, respectively. The state with values $(x_{16}, x_{21}) = (1, 0)$ or $(0, 1)$ is considered to be a desired or undesired state, respectively. Let $\#s^I_{in,x_{16}}$ and $\#s^I_{tar,x_{16}}$ denote the number of states with active x_{16} in s^I_{in} and s^I_{tar}, respectively. When training $FFN^1_{x_{16}}$, we use a lower bound and an upper bound

$lower^I_{x_{16}} = \frac{1}{2}\min\left\{\#s^I_{in,x_{16}}, \#s^I_{tar,x_{16}}\right\}$ and $upper^I_{x_{16}} = 2lower^I_{x_{16}}$ to add the restriction of marker tendency:

$$lower^I_{x_{16}} \le \#FFN^1_{x_{16}}\left(s^I_{in}\right) \le upper^I_{x_{16}}.$$

Similarly, we define $\#s^I_{in,x_{21}}$, $\#s^I_{tar,x_{21}}$, $lower^I_{x_{21}} = \max\left\{\#s^I_{in,x_{21}}, \#s^I_{tar,x_{21}}\right\}$, and $upper^I_{x_{21}} = \frac{3}{2}lower_{x_{21}}{}^I$.

We add the restriction of marker tendency when training $FFN^1_{x_{21}}$:

$$lower^I_{x_{21}} \le \#FFN^1_{x_{21}}\left(s^I_{in}\right) \le upper^I_{x_{21}}.$$

Replacing $\left(\#s^I_{in,x_{16}}, \#s^I_{tar,x_{16}}\right)$ and $\left(\#s^I_{in,x_{21}}, \#s^I_{tar,x_{21}}\right)$ with $\left(\#s^T_{in,x_{16}}, \#s^T_{tar,x_{16}}\right)$ and $\left(\#s^T_{in,x_{21}}, \#s^T_{tar,x_{21}}\right)$, respectively, we add the restrictions of marker tendency when training $FFN^2_{x_{16}}$ and $FFN^2_{x_{21}}$:

$lower^T_{x_{16}} \le \#FFN^2_{x_{16}}\left(s^T_{in}\right) \le upper^T_{x_{16}}$ and $lower^T_{x_{21}} \le \#FFN^2_{x_{21}}\left(s^T_{in}\right) \le upper^T_{x_{21}}$.

Replacing $\left(\#s^I_{in,x_{16}}, \#s^I_{tar,x_{16}}\right)$ with $\left(\#s^F_{in,x_{16}}, \#s^F_{tar,x_{16}}\right)$ and $\left(\#s^{cFFN}_{in,x_{16}}, \#s^{cFFN}_{tar,x_{16}}\right)$, we define

$$\begin{aligned} lower^F_{x_{16}} &= \tfrac{1}{2}\min\left\{\#s^F_{in,x_{16}}, \#s^F_{tar,x_{16}}\right\},\ upper^F_{x_{21}} \\ &= \max\left\{\#s^F_{in,x_{21}}, \#s^F_{tar,x_{21}}\right\}, \\ lower^{cFFN}_{x_{16}} &= \tfrac{1}{2}\min\left\{\#s^{cFFN}_{in,x_{16}}, \#s^{cFFN}_{tar,x_{16}}\right\},\ upper^{cFFN}_{x_{21}} \\ &= \max\left\{\#s^{cFFN}_{in,x_{21}}, \#s^{cFFN}_{tar,x_{21}}\right\} \end{aligned}$$

and add the restrictions when training $FFN^3_{x_{16}}$, $FFN^3_{x_{21}}$, $cFFN_{x_{16}}$, and $cFFN_{x_{21}}$:

$$\begin{aligned} lower^F_{x_{16}} &\le \#FFN^3_{x_{16}}\left(s^C_{in}\right),\quad \#FFN^3_{x_{21}}\left(s^C_{in}\right) \le upper^F_{x_{21}}, \\ lower^{cFFN}_{x_{16}} &\le \#cFFN_{x_{16}}\left(s^I_{in}\right),\quad \#cFFN_{x_{21}}\left(s^I_{in}\right) \le upper^{cFFN}_{x_{21}}. \end{aligned}$$

MAPK33

Nodes x_3 and x_{28} are the markers for desired and undesired state, respectively. The state with values $(x_3, x_{28}) = (1, 0)$ or $(0, 1)$ is considered to be a desired or undesired state, respectively. Replacing numbers (16, 21) of markers (x_{16}, x_{21}) for CACC21 with numbers (3, 28) of markers (x_3, x_{28}) for MAPK33 provides similar restrictions of marker tendency when training iFFN, tFFN, fFFN, and cFFN.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

S-MC and K-HC designed the project, supervised the research, analyzed the results, and co-wrote the manuscript. S-MC and LA performed the modeling and simulation analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Science and ICT (2020R1A2B5B03094920), and the Electronics and Telecommunications Research Institute (ETRI) grant (20ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System).

ACKNOWLEDGMENTS

The authors thank Sea Choi for her critical reading and comments and also thank Hoon-Min Kim for his help in analyzing the measured single-cell data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.594151/full#supplementary-material

REFERENCES

Baslan, T., Kendall, J., Volyanskyy, K., McNamara, K., Cox, H., D'Italia, S., et al. (2020). Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *eLife* 9:e51480.

Berkel, C., and Cacan, E. (2019). Single-cell epigenomics in cancer research. *Biomed. J. Sci. Techn. Res.* 21, 15966–15973.

Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* 19:161.

Chen, H., Albergante, L., Hsu, J. Y., Lareau, C. A., Bosco, G. L., Guan, J., et al. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat. Commun.* 10:1903.

Cho, K.-H., Joo, J. I., Shin, D., Kim, D., and Park, S.-M. (2016). The reverse control of irreversible biological processes. *WIREs Syst. Biol. Med.* 8, 366–377. doi: 10.1002/wsbm.1346

Cho, K.-H., Lee, S., Kim, D., Shin, D., Joo, J. I., and Park, S.-M. (2017). Cancer reversion, a renewed challenge in systems biology. *Curr. Opin. Syst. Biol.* 2, 48–57.

Choi, J., Gong, J.-R., Hwang, C. Y., Joung, C. Y., Lee, S., and Cho, K.-H. (2020). A systems biology approach to identifying a master regulator that can transform the fast growing cellular state to a slowly growing one in early colorectal cancer development model. *Front. Genet.* 11:570546. doi: 10.3389/fgene.2020.570546

Choo, S. M., Park, S. M., and Cho, K. H. (2019). Minimal intervening control of biomolecular networks leading to a desired cellular state. *Sci. Rep.* 9:13124.

Cui, Y., Zheng, Y., Liu, X., Yan, L., Fan, X., Yong, J., et al. (2019). Single-cell transcriptome analysis maps the developmental track of the human heart. *Cell Rep.* 26, 1934–1950. doi: 10.1016/j.celrep.2019.01.079

Duncan, K. D., Fyrestam, J., and Lanekoff, I. (2019). Advances in mass spectrometry based single-cell metabolomics. *Analyst* 144, 782–793. doi: 10.1039/c8an01581c

Grieco, L., Calzone, L., Bernard-Pierrot, I., Radvanyi, F., Kahn-Perles, B., and Thieffry, D. (2013). Integrative modeling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput. Biol.* 9:e1003286. doi: 10.1371/journal.pcbi.1003286

Hamey, F. K., Nestorowa, S., Kinston, S. J., Kent, D. G., Wilson, N. K., and Göttgens, B. (2017). Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5822–5829. doi: 10.1073/pnas.1610609114

He, H., Suryawanshi, H., Morozov, P., Gay-Mimbrera, J., Del Duca, E., Kim, H. J., et al. (2020). Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis. *J. Allergy Clin. Immunol.* 145, 1615–1628. doi: 10.1016/j.jaci.2020.01.042

Huang, P., Zhao, Y., Zhong, J., Zhang, X., Liu, Q., Qiu, X., et al. (2020). Putative regulators for the continuum of erythroid differentiation revealed by single-cell transcriptome of human BM and UCB cells. *Proc. Natl. Acad. Sci. U.S.A.* 117, 12868–12876. doi: 10.1073/pnas.1915085117

Kawai, T., Ota, N., Okada, K., Imasato, A., Owa, Y., Morita, M., et al. (2019). Ultrasensitive Single cell metabolomics by capillary electrophoresis-mass spectrometry with a thin-walled tapered emitter and large-volume dual sample preconcentration. *Anal. Chem.* 91, 10564–10572. doi: 10.1021/acs.analchem.9b01578

Kim, J. R., and Cho, K. H. (2006). The multi-step phosphorelay mechanism of unorthodox two-component systems in *E. coli* realizes ultrasensitivity to stimuli while maintaining robustness to noises. *Comput. Biol. Chem.* 30, 438–444. doi: 10.1016/j.compbiolchem.2006.09.004

Kim, J. R., Kim, J., Kwon, Y. K., Lee, H. Y., Heslop-Harrison, P., and Cho, K. H. (2011). Reduction of complex signaling networks to a representative kernel. *Sci. Signal.* 4:ra35. doi: 10.1126/scisignal.2001390

Kumar, R., Ghosh, M., Kumar, S., and Prasad, M. (2020). Single cell metabolomics: a future tool to unmask cellular heterogeneity and virus-host interaction in context of emerging viral diseases. *Front. Microbiol.* 11:1152. doi: 10.3389/fmicb.2020.01152

Labib, M., and Kelley, S. O. (2020). Single-cell analysis targeting the proteome. *Nat. Rev. Chem.* 4, 143–158. doi: 10.1038/s41570-020-0162-7

Lee, S., Lee, C., Hwang, C.-Y., Kim, D., Han, Y., Hong, S. N., et al. (2020). Network inference analysis identifies SETDB1 as a key regulator for reverting colorectal cancer cells into differentiated normal-like cells. *Mol. Cancer Res.* 18, 118–129. doi: 10.1158/1541-7786.mcr-19-0450

Lu, J., Zeng, H., Liang, Z., Chen, L., Zhang, L., Zhang, H., et al. (2015). Network modeling reveals the mechanism underlying colitis-associated colon cancer and identifies novel combinatorial anti-cancer targets. *Sci. Rep.* 5:14739.

Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., et al. (2019). Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* 176, 1325–1339. doi: 10.1016/j.cell.2019.01.022

Lun, X. K., and Bodenmiller, B. (2020). Profiling cell signaling networks at single-cell resolution. *Mol. Cell. Proteom.* 19, 744–756. doi: 10.1074/mcp.r119.001790

Minakshi, P., Kumar, R., Ghosh, M., Saini, H. M., Ranjan, K., Brar, B., et al. (2019). "Single-Cell proteomics: technology and applications," in *Single-Cell Omics*, eds D. Barh and V. Azevedo (Cambridge, MA: Academic Press), 283–318. doi: 10.1016/b978-0-12-814919-5.00014-2

Schmidt, H., Cho, K. H., and Jacobsen, E. W. (2005). Identification of small scale biochemical networks based on general type system perturbations. *FEBS J.* 272, 2141–2151. doi: 10.1111/j.1742-4658.2005.04605.x

Sreenath, S., Cho, K. H., and Wellstead, P. (2008). Modelling the dynamics of signalling pathways. *Essays Biochem.* 45, 1–28. doi: 10.1042/bse0450001

Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H., et al. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* 146:dev170506. doi: 10.1242/dev.170506

Verma, M., and Kumar, V. (2019). Single-cell epigenomics: technology and applications. *Single Cell Omics* 1, 215–229. doi: 10.1016/b978-0-12-814919-5.00011-7

Yofe, I., Dahan, R., and Amit, I. (2020). Single-cell genomic approaches for developing the next generation of immunotherapies. *Nat. Med.* 26, 171–177. doi: 10.1038/s41591-019-0736-4

Zhu, Y., Scheibinger, M., Ellwanger, D. C., Krey, J. F., Choi, D., Kelly, R. T., et al. (2019). Single-cell proteomics reveals changes in expression during hair-cell development. *eLife* 8:e50777.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fgene.2020.612256

Regulatory Dynamics of Cell Differentiation Revealed by True Time Series From Multinucleate Single Cells

*Anna Pretschner[1], Sophie Pabel[1], Markus Haas[1], Monika Heiner[2] and Wolfgang Marwan[1]**

[1] *Magdeburg Centre for Systems Biology and Institute of Biology, Otto von Guericke University, Magdeburg, Germany,*
[2] *Computer Science Institute, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany*

Dynamics of cell fate decisions are commonly investigated by inferring temporal sequences of gene expression states by assembling snapshots of individual cells where each cell is measured once. Ordering cells according to minimal differences in expression patterns and assuming that differentiation occurs by a sequence of irreversible steps, yields unidirectional, eventually branching Markov chains with a single source node. In an alternative approach, we used multi-nucleate cells to follow gene expression taking true time series. Assembling state machines, each made from single-cell trajectories, gives a network of highly structured Markov chains of states with different source and sink nodes including cycles, revealing essential information on the dynamics of regulatory events. We argue that the obtained networks depict aspects of the Waddington landscape of cell differentiation and characterize them as reachability graphs that provide the basis for the reconstruction of the underlying gene regulatory network.

Keywords: single cell time series, gene regulatory network, Petri net, Markov chain, systems biology, Waddington landscape

Edited by:
Jianhua Xing,
University of Pittsburgh, United States

Reviewed by:
Shou-Wen Wang,
Harvard Medical School,
United States
Junyue Cao,
The Rockefeller University,
United States

***Correspondence:**
Wolfgang Marwan
wolfgang.marwan@ovgu.de

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

***Received:** 30 September 2020*
***Accepted:** 07 December 2020*
***Published:** 08 January 2021*

Citation:
Pretschner A, Pabel S, Haas M, Heiner M and Marwan W (2021) Regulatory Dynamics of Cell Differentiation Revealed by True Time Series From Multinucleate Single Cells. Front. Genet. 11:612256. doi: 10.3389/fgene.2020.612256

INTRODUCTION

Single-cell analyses revealed complex dynamics of gene regulation in differentiating cells (Spiller et al., 2010; Junker and van Oudenaarden, 2014; Paul et al., 2015; Marr et al., 2016; Plass et al., 2018). It is believed that dynamic effects possibly superimposed by stochastic fluctuations in gene expression levels may play crucial roles in cell fate choice, commitment, and reprogramming (Graf and Enver, 2009; Huang et al., 2009; Zhou and Huang, 2011; FerrellJr., 2012; Il Joo et al., 2018; Bornholdt and Kauffman, 2019). Changes in gene expression over time have not been directly measured in single mammalian cells as cells are - for technical reasons - sacrificed during the analysis procedure and hence can be measured only once. Instead, algorithms have been developed to infer the gene expression trajectory of a typical cell in pseudo-time from static snapshots of gene expression states in a cell population, resulting in Markov chains of states (Bendall et al., 2014; Cannoodt et al., 2016; Chen et al., 2019; Saelens et al., 2019; Setty et al., 2019). Most trajectory inference algorithms are based on the assumption that differentiation is unidirectional (Bendall et al., 2014; Haghverdi et al., 2016; Street et al., 2018; Saelens et al., 2019) and that the probability of transiting from one state to the next similar state is independent of the individual history of a cell (Setty et al., 2019). The inference of trajectories has been used to create pseudo-time series for differentiation (Marco et al., 2014; Moignard et al., 2015; Shin et al., 2015; Macaulay et al., 2016), cell cycle (Kafri et al., 2013), and the response to perturbation (Gaublomme Jellert et al., 2015). As any given distribution of expression patterns could result from multiple dynamics, the reconstruction of trajectories from snapshots faces fundamental limits (Weinreb et al., 2018). Even though regulatory mechanisms cannot be directly and rigorously inferred from snapshots (Weinreb et al., 2018), dynamic analyses may be of immediate importance to resolve competing views on basic mechanisms and the role of stochasticity in cell fate decisions (Moris et al., 2016).

True single cell time series can be obtained in *Physarum polycephalum* by taking multiple samples of one and the same giant cell. *Physarum* belongs to the amoebozoa group of organisms. It has a complex, prototypical eukaryote genome (Schaap et al., 2016) and forms different cell types during its life cycle (Alexopoulos and Mims, 1979).

Giant, multi-nucleate cells, so-called plasmodia provide a source of macroscopic amounts of homogeneous protoplasm with a naturally synchronous population of nuclei, which is continually mixed by vigorous shuttle-streaming (Guttes and Guttes, 1961, 1964; Rusch et al., 1966; Dove et al., 1986). The differentiation of a plasmodium into fruiting bodies involves extensive remodeling of signal transduction and transcription factor networks with alterations at the transcriptional, translational, and post-translational level (Glöckner and Marwan, 2017).

In starving plasmodial cells, the formation of fruiting bodies can be experimentally triggered by a brief pulse of far-red light received by phytochrome as photoreceptor (Starostzik and Marwan, 1995b; Lamparter and Marwan, 2001; Schaap et al., 2016). Retrieving small samples of the same plasmodial cell before and at different time points after an inductive light pulse allows to follow how gene expression changes over real time. Because cell cycle, cell fate choice, and development are synchronous throughout the plasmodium (Rusch et al., 1966; Starostzik and Marwan, 1995a; Hoffmann et al., 2012; Walter et al., 2013; Rätzel and Marwan, 2015), single-cell gene expression trajectories can indeed be constructed from time series. By assembling finite state machines made from trajectories we have constructed Petri net models for the state transitions that predict Markov chains as variable developmental routes to differentiation (Werthmann and Marwan, 2017; Rätzel et al., 2020) which may be considered as trajectories through the Waddington landscape (Waddington, 1957; Huang et al., 2009). These Petri nets also predict reversible and irreversible steps, commitment points, and meta-stable states in cells responding to a differentiation stimulus. However, the computational approach for the construction of Petri nets from time series has been originally developed with data sets of a coarse resolution in time and the structural resolution of the nets was accordingly limited. Nevertheless, the approach turned out to be useful for capturing the dynamics of the process. For this paper, we developed a method for retrieving smaller samples from even larger plasmodial cells and showed that these cells provide a homogeneous source for samples to be taken. This allowed us to considerably improve the time resolution as compared to previous studies. Sampling cells at higher time resolution, allowed the construction of Petri nets with enhanced structural and dynamic resolution. Structural complexity, highly connected nodes, parallel pathways, reversible reactions, and Petri net places representing meta-stable states in the developmental network, as revealed by the new data sets, characterize the differentiation response as complex and dynamic in contrast to a smooth, continuous process. We describe the graph properties of the Waddington Petri nets and conclude that the gene expression dynamics revealed by our analysis most likely emerge from the non-linear dynamic behavior of the underlying regulatory network rather than from stochastic fluctuations in the concentration of regulatory molecules.

MATERIALS AND METHODS

Plasmodial Strain, Growth of Cells, Sample Preparation, and Gene Expression Analysis

Sporulation-competent plasmodial cells of wild type strain LU897 × LU898 (Starostzik and Marwan, 1998) were obtained as previously described (Starostzik and Marwan, 1998; Rätzel et al., 2020). A total of 2.8 gram of plasmodial mass was applied to a 14 cm Ø Petri dish that contained 90 ml of semi-rich Golderer agar (Golderer et al., 2001), based on a salt solution of 0.01% (w/v) niacin, 0.01% (w/v) niacinamide, 0.1% (w/v) $CaCO_3$, and 0.14 mM $CuCl_2$, supplemented with 5 g peptone from meat (Sigma Aldrich), 0.75 g yeast extract (Becton, Dickinson & Co.), and 3.9 mM glucose per liter, adjusted to pH 4.6 with concentrated HCl. After starvation for 7 days at 22°C in complete darkness, sporulation was induced with a 15 min pulse of far-red light ($\lambda \geq 700$ nm, 13 W/m^2) (Starostzik and Marwan, 1998). Before and at 1-h time intervals after the start of the far-red pulse, samples were taken in duplicate at arbitrarily chosen but distant positions on the plate. Each sample was obtained by picking an agar plug of 1.13 cm^2 with the cut bulb of a disposable Pasteur pipette (EA62.1; Carl Roth, Karlsruhe, Germany). The plasmodial mass on the agar plug was scraped off with a pipet tip and, by cutting the tip, transferred into a vial of glass beads immersed in liquid nitrogen (**Figure 1**). After extraction of RNA and removal of contaminating DNA (Marquardt et al., 2017), the relative abundance of the mRNAs of 35 genes, differentiation marker and reference genes (Hoffmann et al., 2012; **Supplementary Table 2**) was analyzed by gene expression profiling (GeXP), a multiplex RT-PCR method (Hayashi et al., 2007) as previously described (Rätzel and Marwan, 2015; Marquardt et al., 2017).

Data Analysis Pipeline and Automated Generation of Petri Nets

To correct for differences in the concentration of total RNA and in the efficiency of the RT-PCR reaction, the gene expression values were normalized to the median of the estimated relative concentrations of mRNAs of the 35 genes in each RNA sample. Each normalized expression value was subsequently normalized

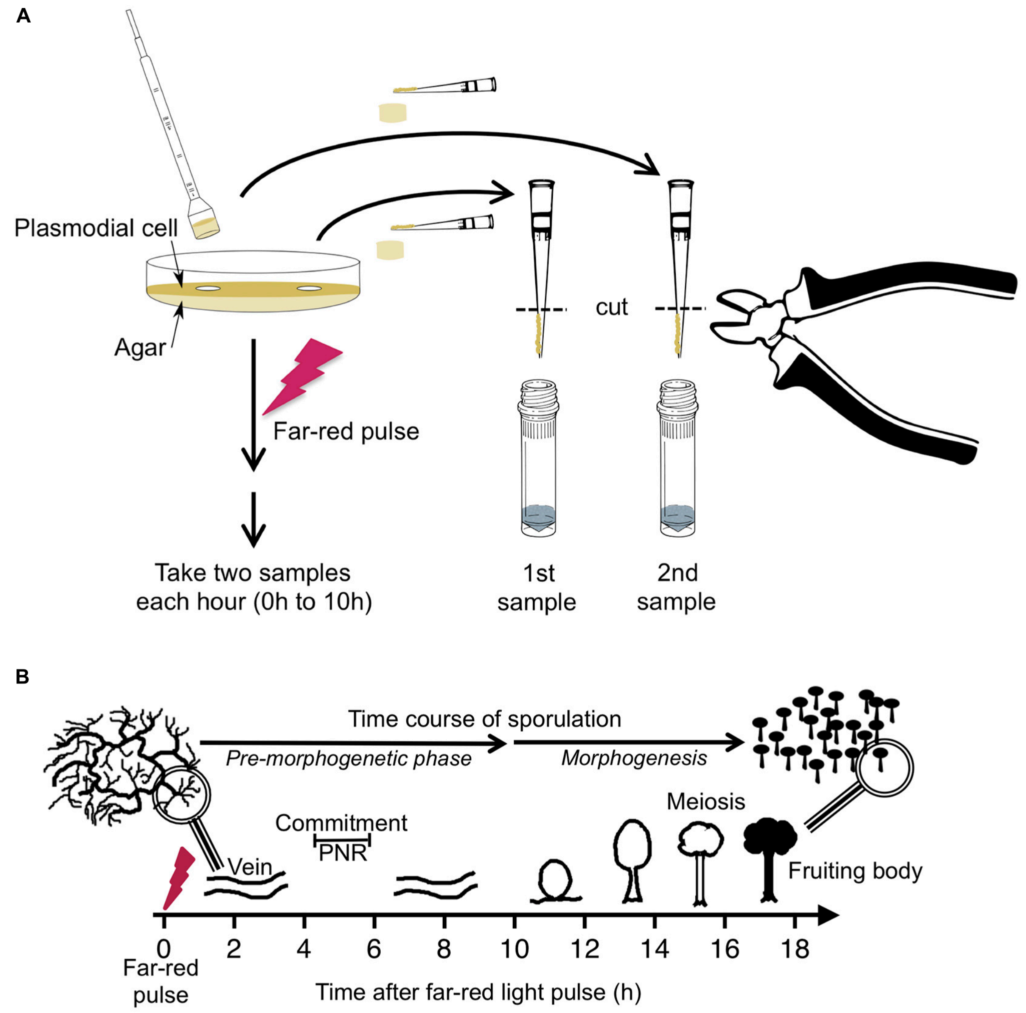


FIGURE 1 | Experimental protocol for taking time series by repeated sampling of individual plasmodial cells and time course of light-induced sporulation. **(A)** Each Petri dish contained one individual plasmodial cell supported by an agar substratum. Before and at 1-h time intervals after stimulation of the cell with a pulse of far-red light, samples were taken in duplicate by picking an agar plug at arbitrarily chosen but distant positions on the plate. The cell mass was scraped off from the agar plug with a pipet tip and transferred into a vial containing glass beads and liquid nitrogen. After purification of RNA, the gene expression pattern was estimated twice in each sample, each with two independent multiplex RT-PCR reactions (see section "Materials and Methods" for details). For dark controls, the far-red light stimulus was omitted. **(B)** Time frame of light-induced sporulation of a plasmodial cell. At about four to 6 h after stimulation with a pulse of far-red light, the cell is irreversibly committed to sporulation by crossing the point of no return (PNR) while there is no obvious change in the plasmodial morphology. Morphogenesis then starts at about 11 h after the stimulus by the formation of nodules that subsequently culminate to form the fruiting bodies. Panel B was taken from Glöckner and Marwan (2017).

to the geometric mean of all values obtained for a given gene, and this was performed separately for each gene.

Data were analyzed and processed with a revised and extended pipeline written in R (R Core Team., 2016), based on the previously described script (Rätzel et al., 2020). The normalized gene expression data were clustered and significant clusters were determined with the help of the Simprof algorithm (Clarke et al., 2008) as provided by the clustsig package (Whitaker and Christman, 2014). Expression patterns were visualized in the form of a heatmap generated by the heatmap.2 function, provided as part of the gplots package (Warnes et al., 2016). Changes in gene expression over time were visualized by multidimensional scaling based on Euclidean distance (Gower, 1966) with the help of the *cmdscale* function provided as part of the stats package v3.5.1 (R Core Team., 2016). Petri nets were constructed from single cell trajectories of gene expression as previously described (Rätzel et al., 2020). Each trajectory is a temporal sequence of gene expression states, where each state corresponds to a Simprof significant cluster. Petri nets specified in ANDL format (Abstract Net Description Language) (Heiner et al., 2013) were imported into Snoopy (Rohr et al., 2010) and graphically displayed by running the Sugiyama layout algorithm (Sugiyama et al., 1981). Petri net places, each representing a gene expression state, were colored according to the relative temporal stability of the expression state or according to the relative frequency with which each gene expression state occurred. Petri net transitions, corresponding to transits between states were colored according to the frequency with which each transit occurred or according to the data subset in which the transit occurred. These parameters were computed and coloring was performed by automatic editing of Snoopy files encoded in xml format, again with the help of a R script. The raw data and the complete computational pipeline used in this study is provided as part of the **Supplementary Material**.

RESULTS

Even Large Plasmodia Provide a Source of Homogeneous Cell Material for True Time Series Analysis of Gene Expression

In previous studies we have shown that the gene expression pattern in samples taken at the same time from different sites of a plasmodium covering a standard Petri dish (9 cm Ø) did not change within the limits of accuracy of the measurements. Accordingly, repeated sampling of the same plasmodial cell yields true time series (Rätzel, 2015; Werthmann and Marwan, 2017). To allow more samples to be taken without consuming too much of the plasmodial mass, we now prepared plasmodia on 14 cm Ø Petri dishes, increasing the surface area covered by the plasmodial mass by 2.4-fold, and took smaller samples by punching agar plugs of 1.13 cm^2 per sample, to harvest a small portion of the initial total plasmodial mass. To test whether the homogeneity in gene expression is impaired or even lost in the larger plasmodia, we took 9 or 16 samples at the same time from approximately evenly spread sites of a plasmodium, and estimated the gene expression pattern twice in the RNA of each sample, to obtain one technical replicate of each measurement (Aselmeyer, 2019; Driesch, 2019). This allowed to estimate the biological variation in gene expression within a plasmodium as compared to the technical accuracy of the measurements. In order to correct for potential differences in the efficiency of the RT-PCR between reactions, the expression value for each gene was normalized to the median of the expression values of all genes measured in the sample (for details see Materials and Methods). To estimate the technical accuracy of the measurements, the relative deviation of first and second measurement from the mean of the two measurements was estimated for each assayed gene in each of the retrieved plasmodial samples. The frequency distribution of all values was almost symmetrical with a tail consisting of a small number of low values, obviously as a result of inefficient RT-PCR reactions. To estimate the degree of homogeneity in gene expression within a plasmodial cell, we asked to which extent the expression values for the 9 or 16 samples taken from the same plasmodium deviated from their median. To restrict the influence of technical artifacts on the result, we considered the subset of the data where first and the second measurement of the same plasmodial sample deviated not more than two-fold from the mean of the two values. In three of the total of 46 analyzed plasmodia (30 far-red stimulated; 16 dark controls), individual samples deviated from the rest of the samples of the same plasmodium by more than a factor of two. As errors in sample preparation could not be ruled out, these three plasmodia were excluded and the remaining data set of 43 plasmodia (28 far-red stimulated; 15 dark controls) was analyzed taking the mean of 1st and 2nd measurement for each gene in each sample. Among the total of 7160 values, 98% of the symmetric frequency distribution (**Supplementary Figure 1**) were between 0.48- and 2.10-fold deviation of the median of all values of the respective plasmodium (**Supplementary Table 1**). There was no obvious difference between dark controls and far-red stimulated plasmodia which were measured at 6 h after the light pulse when genes were already differentially regulated (**Supplementary Figure 1** and **Supplementary Table 1**), indicating that even during the period where the mRNA abundance changed in time, the homogeneity in gene expression levels is maintained. Visual inspection of outliers within the distribution did not reveal any candidates for specific genes that might be inhomogeneously expressed.

In summary, the gene expression values throughout a plasmodium deviated not more than approximately two-fold from the median of all samples from the same plasmodium and were thus within the limits of the technical accuracy of the measurements, even under conditions were genes were in the process of being up- or down-regulated. These differences measured between samples were minor as compared to the differential regulation where the expression level of genes changed in the order of ten to more than hundred-fold (**Supplementary Figure 2**). These results are consistent with the results of the time-series experiment, where for each time point, two samples were retrieved and analyzed from the same plasmodial cell (see below).

Sampling of Plasmodia at 1 h Time Interval

As the assayed genes were evenly expressed and changed evenly in time throughout the large plasmodia, at least within the limits of accuracy of the measurements, we took time series at 1 h time intervals. In order to assay, in each experiment, for the homogeneity and synchrony in gene expression throughout the plasmodium, we took two samples at each time point from different, arbitrarily chosen but distant sites of the plasmodium (**Figure 1**). In far-red stimulated plasmodia, the first samples (referred to as the 0 h samples) were taken at the start of the experiment, i.e., immediately before application of the 15 min pulse of far-red light. All subsequent samples were taken at 1 h time intervals until 10 h after the start of the experiment [At 5 to 6 h after the far-red pulse cells have passed the commitment point, while visible morphogenesis starts several hours later by entering the transient nodulation stage at about 11 h after the pulse (Hoffmann et al., 2012)]. In the dark controls, the far-red stimulus was omitted. Gene expression in each plasmodial sample taken at a given time point was analyzed twice by GeXP-RT-PCR, where the measurement and the corresponding technical replicate are referred to as 1st and 2nd measurement for sample #1, and 3rd and 4th measurement for sample #2, respectively. Data were normalized as described above.

The technical quality of the measurements was estimated separately for the two data sets, each comprising the data of the samples collected at the 11 time points of the time series. For each plasmodial sample, the relative deviation of the two measurements (1st and 2nd, or 3rd and 4th) from the mean of the two measurements was estimated. The frequency distributions of the deviations and corresponding quantil values indicated that the technical qualities of 1st and 2nd, as well as 3rd and 4th measurement were virtually identical with 95% of the values differing less than a factor of two from each other (**Figure 2** and **Table 1**).

The degree of spatial variability of gene expression within a plasmodium was estimated by combining the data sets for the first and the second sample of a plasmodium taken at each time point of the time series. The frequency distribution of the deviation of each measurement from the mean of 1st, 2nd, 3rd, and 4th measurement of the two samples taken from each plasmodium at any time point was virtually identical to the frequency distributions obtained for the technical replicates, indicating that gene expression within the analyzed plasmodia varied at maximum within the limits of accuracy of the measurements (within a factor of 2 in 95% of the samples). This conclusion is based on the comparison of the quantile distributions of the data sets (**Figure 2** and **Table 1**) considering a total of 36,540 data points.

Multi-Dimensional Scaling Analysis

With this data set, we investigated how expression changes as a function of time in the individual plasmodial cells. The gene expression pattern of a plasmodial cell at a given time point was obtained as the mean of the four expression values of each gene measured in the two plasmodial samples picked at that time point.

For visual representation of the data set and of single-cell trajectories of gene expression, we performed multidimensional scaling (MDS) to obtain a data point for the expression pattern of each cell at each time point. Single-cell trajectories of gene expression are shown in **Figure 3**. Notably, the gene expression patterns of un-stimulated cells (dark controls) changed as a function of time with the highest variability along coordinate 2 of the MDS plot **Figures 3A,B**. Trajectories of far-red stimulated cells (**Figures 3C,D**) moved from the left side to the right side of the plot, while the shape of individual trajectories varied to a certain extent, indicating that the response of the cells was similar though not identical. Obviously, the trajectories of six of the eight far-red-stimulated plasmodia of experiment #1 traversed a considerably larger area of the MDS plot (**Figure 3C**) as compared to the other stimulated cells, indicating a larger variation in gene expression during the response to the stimulus. The extent of variation is accordingly obvious when the bulk of data points is placed in the same plot (**Figure 4A**). To search for genes that may account for the scattering along coordinate 2, we visually inspected the individual time series displayed in the form of a heat map (**Supplementary Figure 2**). In addition to the genes that were clearly up- or down-regulated in response to the stimulus, the messages of four genes, *hstA*, *nhpA*, *pcnA*, and *uchA*,

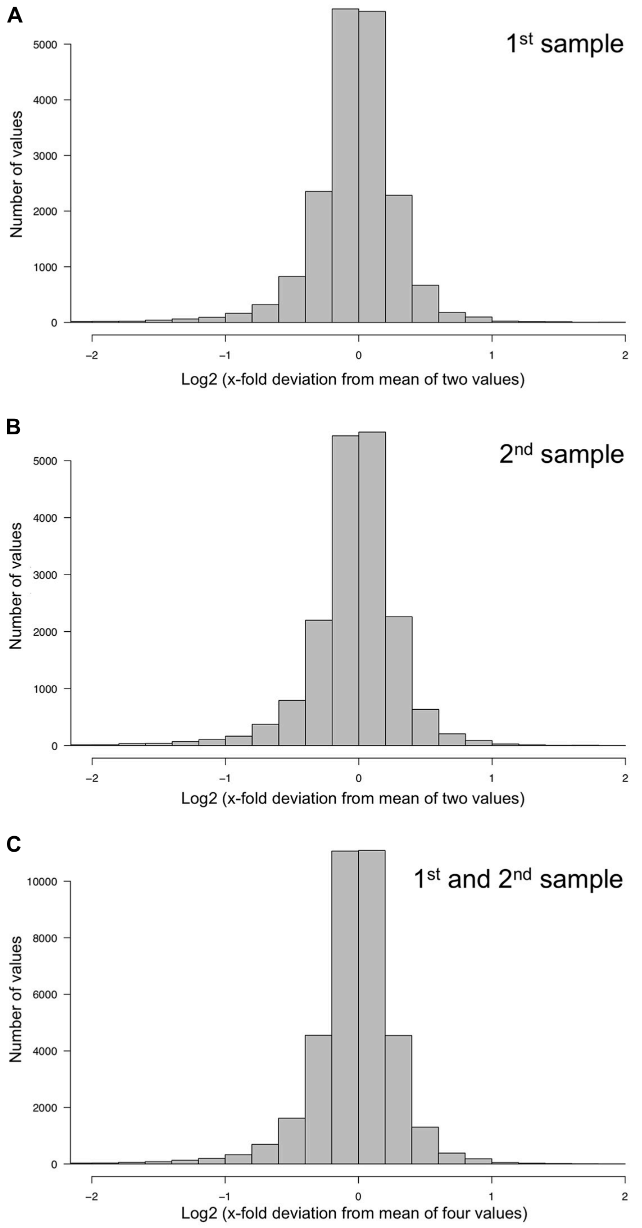


FIGURE 2 | Technical accuracy of measurements and homogeneity of gene expression within a plasmodial cell as determined by technical and biological replicates, taken in experiments #1 and #2 (**Table 2**). **(A,B)** Technical accuracy of measurements of gene expression. The concentration of the mRNAs of the set of 35 genes (**Supplementary Table 2**) was determined twice by RT-PCR for each RNA sample. The frequency distributions display the Log2 of the x-fold deviation of each expression value of each gene from the mean of the two values obtained by technical replication. Panels **(A,B)** show the results obtained for each of the two biological samples [**(A)**, sample #1; **(B)**, sample #2], that both were simultaneously taken from the same plasmodial cell at any time point during the experiments. **(C)** Combination of the data sets shown in panels **(A,B)**. This frequency distribution shows the deviation of each measurement from the mean of four values, obtained by twice measuring each of the two biological samples simultaneously taken from the same plasmodium at any time point of the experiments. The figure represents the complete data set of 36,540 data points that was analyzed in the present study.

TABLE 1 | Quantil distributions of the x-fold deviation (x) of a value from the mean of two or four values, characterizing the reproducibility of measurements as estimated through technical and biological replicates, respectively.

Measurements	1st and 2nd		3rd and 4th		1st to 4th	
Percent of values	**Quantile (Log2 (x))**	**Quantile (x)**	**Quantile (Log2 (x))**	**Quantile (x)**	**Quantile (Log2 (x))**	**Quantile (x)**
1%	−1.308	0.404	−1.383	0.383	−1.349	0.392
5%	−0.555	0.681	−0.592	0.663	−0.574	0.672
25%	−0.169	0.890	−0.169	0.890	−0.169	0.890
50%	−0.010	0.993	−0.008	0.995	−0.009	0.994
75%	0.137	1.100	0.142	1.103	0.139	1.101
95%	0.414	1.333	0.415	1.334	0.415	1.333
99%	0.750	1.682	0.746	1.677	0.747	1.679

*The table quantitatively characterizes the frequency distributions shown in **Figure 2**.*

in the following called *pcnA*-group genes, changed over time in some of the plasmodia, but there was no obvious consistent relationship to the time point of stimulus application. When only genes were included in the analysis that were clearly up- or down-regulated in response to the stimulus (**Supplementary Figure 2**, see also **Figure 8**), the data points of the MDS plot were indeed less scattered (**Figure 4B**). A qualitatively similar result was obtained plotting the up-regulated and the down-regulated genes separately (**Figures 4C,D**), with some more variation in the expression of the down-regulated genes. However, expression of the *pcnA*-group over time (**Figure 4E**) was clearly different from the up- or down-regulated genes. Expression of the pcnA-group genes was different between cells from experiments #1 and #2 as seen from the trajectories of the cells (**Supplementary Figure 3**), suggesting that ongoing internal processes in cells of experiment #1 might even influence their response to the light stimulus. Indeed, according to the corresponding MDS plots of the bulk data points (**Supplementary Figure 4**) the response of the up- and down-regulated genes was less uniformly in cells of experiment #1 (**Supplementary Figure 4C**) as compared to those of experiment #2 (**Supplementary Figure 4D**).

Construction and Graph Properties of Waddington Landscape Petri Nets

For a further analysis, we performed hierarchical clustering of the expression data for all assayed 35 genes, differentiation marker and reference genes (**Supplementary Table 2**), and identified significantly different clusters of expression patterns with the help of the Simprof algorithm (Clarke et al., 2008; **Supplementary Figure 5**). The temporal sequences of gene expression patterns classified as Simprof significant clusters defined a trajectory for each individual cell and revealed significant differences between cell trajectories (**Table 2**). To relate gene expression states and trajectories we constructed a Petri net (bipartite graph) as previously described (Werthmann and Marwan, 2017; Rätzel et al., 2020), by representing each gene expression state by a *place* and the temporal transit between two states by a *transition* (**Figure 5**). A single token marking one place of the Petri net indicates the current gene expression state of a cell. The token moves from its place to a downstream place when the transition, connecting the two places through directed arcs, fires. As each transition is connected to exactly two places (one pre-place and one post-place), tokens are neither formed nor destroyed when moving through the net, so the gene expression state of the cell remains unequivocally defined at any time. The coherent Petri net obtained this way represents a state machine predicting possible developmental trajectories in terms of Markov chains of gene expression states (Rätzel et al., 2020).

The basic modeling principles are summarized in **Table 3**. We observe the following structural properties of the model which we call '*Waddington landscape Petri net*':

- Each transition has exactly one pre-place and one post-place.
- There are places having more than one post-transition. These post-transitions are in conflict. But, because every transition has exactly one preplace, each conflict is a free choice conflict, meaning the token is free to choose which route to take, predicting a corresponding free choice for the cell (see Discussion).
- There are places having more than one pre-transition, i.e., alternative paths may re-join. Thus, the Petri net structure does not form a tree.
- There are cycles: a cell may switch back to previous states or oscillate between states as defined by the expression patterns of the set of observed genes.

For technical reasons we add immediate transitions starting alternative trajectories, in order to get a statistical distribution of states in which the experiments have started or will start with a given probability.

In contrast to most state-of-the-art pseudo-time series approaches found in the literature (Saelens et al., 2019), the structure of the Waddington landscape Petri net is not restricted to a partial order, meaning it is neither restricted to a directed acyclic graph nor to a tree. Instead we obtain what is known in Petri net theory as '*state machine,*' also called in other communities 'finite state machine' or 'finite automata,' which may involve cycles.

A state machine with one token and its reachability graph, or Markov chain for stochastic Petri nets, are isomorph (i.e., have the same structure, there is a 1-to-1 correspondence); to put it differently: our (stochastic) Petri net represents the

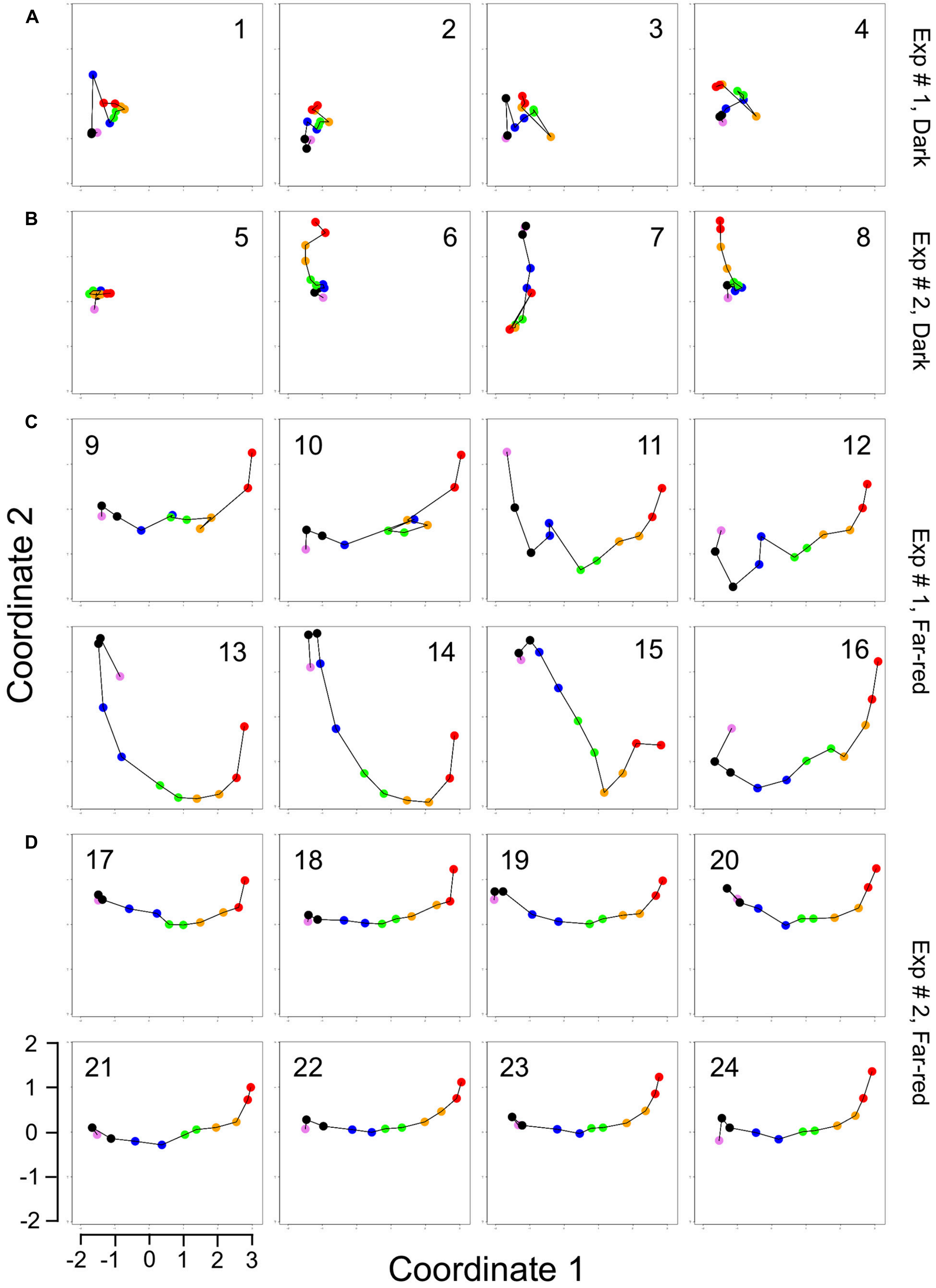


FIGURE 3 | Single cell trajectories of gene expression displayed after multidimensional scaling (MDS) of the expression patterns. Panels **(A,B)** show the trajectories of unstimulated cells of experiment #1 and experiment # 2, respectively. Panels **(C,D)** show the trajectories of far-red stimulated cells of the two experiments. Each data point represents the gene expression pattern of a cell at a given time point at 1 h time intervals. The start position (0 h) of each trajectory is encoded in pink and the endpoint (10 h) in red. The number displayed in each subpanel refers to the ID number of the plasmodial cell (P1 to P24) as listed in **Table 2**. All cells of the dark controls did not sporulate while all far-red irradiated cells sporulated. All plots are displayed at the same scale.

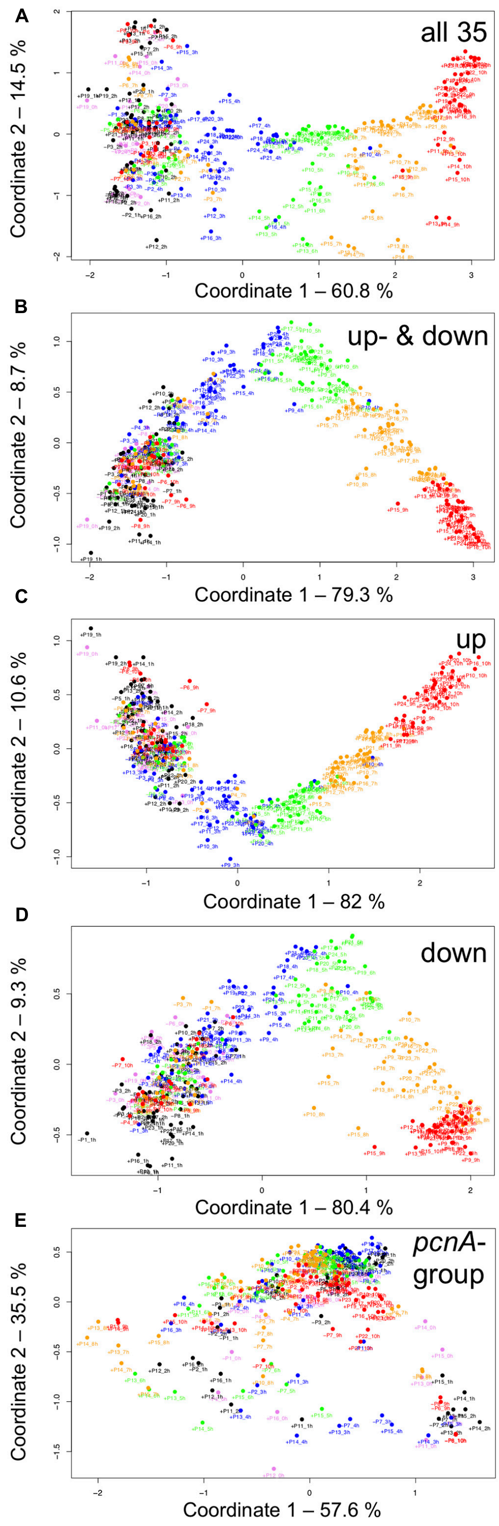


FIGURE 4 | Gene expression patterns of all analyzed cells displayed for different sub sets of genes. Multidimensional scaling was performed for the complete set of 35 genes **(A)**, the subset of the up- and the down-regulated genes **(B)**, or exclusively for up-regulated **(C)**, down-regulated **(D)** or the *pcnA*-group of genes **(E)**. Each data point represents the expression pattern *(Continued)*

FIGURE 4 | Continued
of an individual cell at a given time point. Time is encoded by color (0 h, pink; 1 h, 2 h, black; 3 h, 4 h, blue; 5 h, 6 h, green; 7 h, 8 h, ocher; 9 h, 10 h, red). Developmental destiny, cell number (as assigned in **Table 2**), and time is given for each data point. The label + P12_1 h, for example, indicates that the data point refers to the expression pattern of plasmodium number 12 as measured at 1 h after the start of the experiment (corresponding to the onset of the far-red light stimulus in light-stimulated cells) and that the plasmodium had sporulated (+) in response to the stimulus (+, sporulated; -, not sporulated). The percent of variance is given for each coordinate.

Markov chain of states the cells assume in the course of their developmental trajectory and accordingly on their walk through the Waddington landscape. We assume that the Petri net represents the corresponding region of the Waddington landscape predicting possible developmental paths a single cell can follow, which of course yields a state machine.

Representing Markov chains as Petri nets comes with a couple of advantages.

First, Petri nets are equipped with the concept of T-invariants, which belong to the standard body of Petri net theory from very early on (Lautenbach, 1973). We consider T-invariants as crucial in terms of biological interpretation of the generated net structures (Sackmann et al., 2006; Heiner, 2009). The computation of T-invariants is rather straightforward for state machines; due to their simple structure it holds:

- each cycle in a state machine defines a T-invariant, and
- each elementary cycle (no repetition of transitions) is a minimal T-invariant.

Second, modeling the differentiation-inducing stimuli, what we have not done so far, would turn some of the free choice conflicts into non-free choice conflicts, which involves, technically speaking, leaving the state machine net class. To unequivocally identify transits that are stimulus-dependent, we need a higher data density which we will hopefully achieve in one of our next experiments. With stimulus-dependent transitions, the constructed Petri nets and their Markov chains do not coincide anymore, instead the Markov chains as well as the reachability graph are directly derived from the Petri nets and may be analyzed by standard algorithms. Finally, our Petri net approach paves the way for the actual ultimate goal of our future work - reconstructing the underlying gene regulatory networks based on the reachability graphs encoded by the Waddington landscape Petri nets.

Characterization of Petri Nets Constructed From Gene Expression Data

Figure 6 displays a Petri net assembled from cell trajectories considering the set of 35 genes. The graphical representation laid out using the Sugiyama algorithm emphasizes the directionality of concurrent processes (Sugiyama et al., 1981). The net indicates that cells started in different states (connected to C0; see Legend to **Figure 5** for details) and, after stimulation by far-red light, proceeded to a small set of terminal states (places C71, C76,

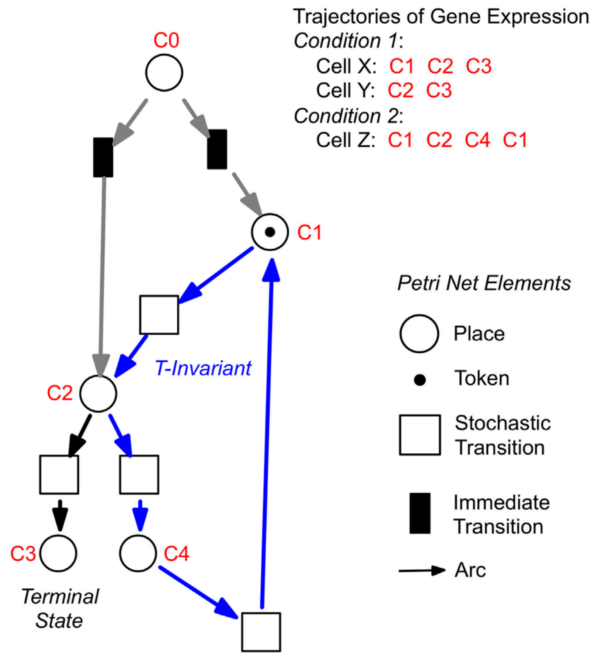


FIGURE 5 | Construction of Petri nets from single cell trajectories of gene expression. Petri nets are directed, bipartite graphs with two types of nodes, places and transitions, that are connected by arcs (see symbols, lower right). Petri nets are used in this work to model state machines, as exemplified in the following. Any gene expression state of a cell, as defined by its assignment to a Simprof significant cluster of gene expression patterns, is represented by a corresponding place (drawn as a circle). Any transit between two states is mediated by a transition (drawn as a rectangle). The current gene expression state of a cell is indicated by one token which marks the respective place. When a place contains a token, one of its post-transitions can fire to move the token into its post-place. (A post-transition of a place is a downstream transition which is immediately connected to that place, as indicated by a directed arc). Because each transition of the Petri nets as they are used here, has exactly one pre-place (one incoming arc) and one post-place (one outgoing arc), and because all arc weights are one, tokens can neither be produced nor destroyed, and the state of gene expression remains unequivocally defined. A transit cycle, i.e., the ensemble of reactions that bring a subsystem back to the state from which it started, is called transition-invariant (T-invariant) (Sackmann et al., 2006). The arcs that contribute to the T-invariant of the Petri net displayed in the figure are highlighted in blue. Petri nets, as they are used in this paper, contain one additional place C0, which does not represent a gene expression state. For simulation, C0 defines the initial gene expression state of the cell by randomly delivering its token to one of the places that are connected to C0 through so-called *immediate transitions* (filled in black) that fire immediately when the simulation starts [for details see (Rätzel et al., 2020)]. Connection to C0 also graphically highlights the places representing those gene expression states in which cell trajectories started. In the example shown, the cell trajectory started in a gene expression state assigned to Simprof cluster C1. The token can move to C2 where it randomly moves to either C3 (a terminal state in this example) or to C4, from which it may return to C1 and possibly continue.

C77, C78, C79, C95) via multiple, more or less highly connected intermediate states. To facilitate the interpretation of the Petri net, we have colored the places and transitions according to different criteria. In **Figure 6A**, the transitions being specific to cells of experiments #1 or #2 and for dark controls or light-stimulated cells in the respective experiments, are colored differently. Each place is colored according to the relative frequency of its corresponding gene expression state, indicating that some states occurred more frequently than others. From this representation it is obvious that cells of experiments #1 and #2 form different branches, in part projecting onto different terminal states, reflecting accordingly different developmental trajectories to commitment and sporulation. **Figure 6B** displays the same Petri net, but with a different color coding. Here, transitions are colored according to how frequent the corresponding transits occurred, indicating that some paths were more frequently taken than others. Places are colored according to their relative stability, defined as the average residence time of a cell in the respective state (see Methods for details). Coloring indicates that cells reached a terminal state through states of different stability, e.g., meta-stable intermediates.

Un-stimulated cells (**Table 2**, dark controls) spontaneously switched between significantly different states of gene expression. Their trajectories gave three disconnected Petri nets (**Figure 7A**; the three nets were connected to C0 for technical reasons, see legend to **Figure 5**).

Petri nets of **Figures 6A,B, 7A** indicate that the expression pattern in both, stimulated and un-stimulated cells developed predominantly in forward directions while there were some transits back to previous states creating so-called transition-invariants (T-Invariants; **Figure 5**). We asked whether stimulus-independent temporal expression differences like those observed in the subset of the *pcnA*-group of genes might have added to this directedness. Therefore, we constructed a Petri net from trajectories based on significant clusters, this time exclusively clustering the subset of up- and down-regulated genes. Basic features found in the Petri nets of up- and down-regulated genes were similar to the ones found in the Petri nets for the full set of 35 genes: Trajectories formed parallel main branches, there were intermediate nodes of different stability, of different connectedness, and hence states that occurred with different frequency (**Supplementary Figure 6**). In contrast, there was a high number of minimal T-invariants that heavily involved places representing gene expression states that occurred in un-stimulated cells. A Petri net built by considering only transits that occurred in un-stimulated cells (**Figure 7B**) was nearly covered with T-Invariants, indicating spontaneous, reversible alterations in the expression of up- and/or down-regulated genes. Again, states of gene expression displayed different stability. Considerable variation in the expression level of the up-and down-regulated genes is even most obvious from the heat map of initial states from which trajectories emerged and of terminal states that were observed during the experiment (**Figure 8**). The high density of T-invariants in the un-stimulated cells suggests that similarly, the T-invariants involving places corresponding to light-stimulated cells are due to gene expression changes that do spontaneously occur before cells are caught by a new attractor formed in response to the far-red stimulus (see Discussion).

Figure 7 also shows that selecting sets or subsets of genes for hierarchical clustering and subsequent Petri net construction may yield Petri nets of different structure delivering accordingly non-redundant information on corresponding subsets. This is also shown in **Table 4** for the set of 35 genes and for subsets, the down-regulated, up-regulated, down- and up-regulated, and the *pcnA*-group of genes. Except for the *pcnA*-group, the number of places per gene was approximately the same. The number of transitions per gene however became less with more genes considered. This suggests that up-regulation, down-regulation and even expression of the *pcnA*-group of genes are at least partly coordinated or co-regulated processes. The average number of minimal T-invariants per gene compared for the different groups of genes (**Table 4**) suggests that reversibility observed for the subsets vanishes when more genes are considered, obviously due to combinatorial effects.

Single Cell Trajectories Reveal Qualitatively Different Patterns in Differential Gene Regulation

We have argued that Petri net modeling disentangles the complex response of cell reprogramming (Rätzel et al., 2020) and predicts feasible developmental pathways through the Waddington landscape, resulting in significantly distinct single cell trajectories. To reveal similarities and differences of the expression kinetics of individual genes, we plot the geometric mean of the concentration values of the mRNA in a cluster

TABLE 2 | Single cell trajectories of gene expression.

Experiment	Cell	0.0 h	1.0 h	2.0 h	3.0 h	4.0 h	5.0 h	6.0 h	7.0 h	8.0 h	9.0 h	10.0 h
Exp #1, Dark	P1	60	56	59	26	35	34	34	28	28	32	31
Exp #1, Dark	P2	57	57	60	63	35	35	35	14	32	32	32
Exp #1, Dark	P3	59	56	26	36	34	34	28	13	31	32	32
Exp #1, Dark	P4	37	37	37	33	28	28	28	15	29	30	30
Exp #2, Dark	P5	24	52	52	52	49	52	52	52	48	47	48
Exp #2, Dark	P6	25	45	45	44	44	43	43	21	7	5	1
Exp #2, Dark	P7	3	3	3	11	12	63	63	60	60	26	61
Exp #2, Dark	P8	46	45	45	42	42	42	43	21	7	6	1
Exp #1, Far-red	P9	33	23	27	16	101	111	105	87	88	70	79
Exp #1, Far-red	P10	36	23	27	16	88	111	105	89	80	75	78
Exp #1, Far-red	P11	10	53	54	54	15	98	100	89	91	93	71
Exp #1, Far-red	P12	53	55	54	54	15	100	100	89	91	92	71
Exp #1, Far-red	P13	8	4	4	12	53	69	69	68	66	65	95
Exp #1, Far-red	P14	22	4	2	9	53	69	69	68	66	65	95
Exp #1, Far-red	P15	22	4	2	9	96	96	69	67	64	94	95
Exp #1, Far-red	P16	62	55	58	54	97	99	89	90	93	75	78
Exp #2, Far-red	P17	50	40	45	19	104	103	105	87	82	72	79
Exp #2, Far-red	P18	51	41	25	19	104	102	110	85	81	72	77
Exp #2, Far-red	P19	20	20	20	17	18	102	109	83	82	72	77
Exp #2, Far-red	P20	44	40	45	19	103	108	110	84	82	73	77
Exp #2, Far-red	P21	51	39	45	19	103	106	110	84	82	74	76
Exp #2, Far-red	P22	52	39	45	19	103	107	110	86	81	73	76
Exp #2, Far-red	P23	46	39	45	19	103	107	110	86	81	73	77
Exp #2, Far-red	P24	24	38	45	19	104	108	110	84	81	74	77

Trajectories are displayed as temporal sequences of gene expression states. Each state is given by the cluster ID number to which it was assigned by the Simprof algorithm. The two experiments, Exp #1 and Exp #2, were performed on two different days, respectively, with the same strain (LU897 × LU898) and under virtually identical experimental conditions. All cells of the dark controls did not sporulate while all far-red irradiated cells sporulated.

TABLE 3 | Terminology, basic modeling principles, and Petri net elements.

Place	Cellular state as defined by a significant cluster of gene expression patterns.
Transition	Transit, i.e., reprogramming step as defined by a discrete change in the state of gene expression.
Source place (no predecessor node)	The cellular state in which an experiment (recording of a time series) starts. There can be various source nodes, depending on the particular state in which a cell is in the moment when the experiment starts.
Sink node (no successor node)	Place with no outgoing arc, indicating a terminal state of gene expression reached at the end of the experiment. There can be multiple sink nodes.
Conflict	Forward branching places modeling bifurcation.
Token	The token indicates the cell and its current gene expression state; there is always just a single token.
Path	A single path from a source node to a sink node represents a possible developmental trajectory of a single cell. But a trajectory may not necessarily involve a source node and/or a sink node.
Petri net	The entire Petri net gives, for the genes analyzed, that part of the Waddington landscape through which cells passed and accordingly all corresponding developmental trajectories a single cell may undergo.

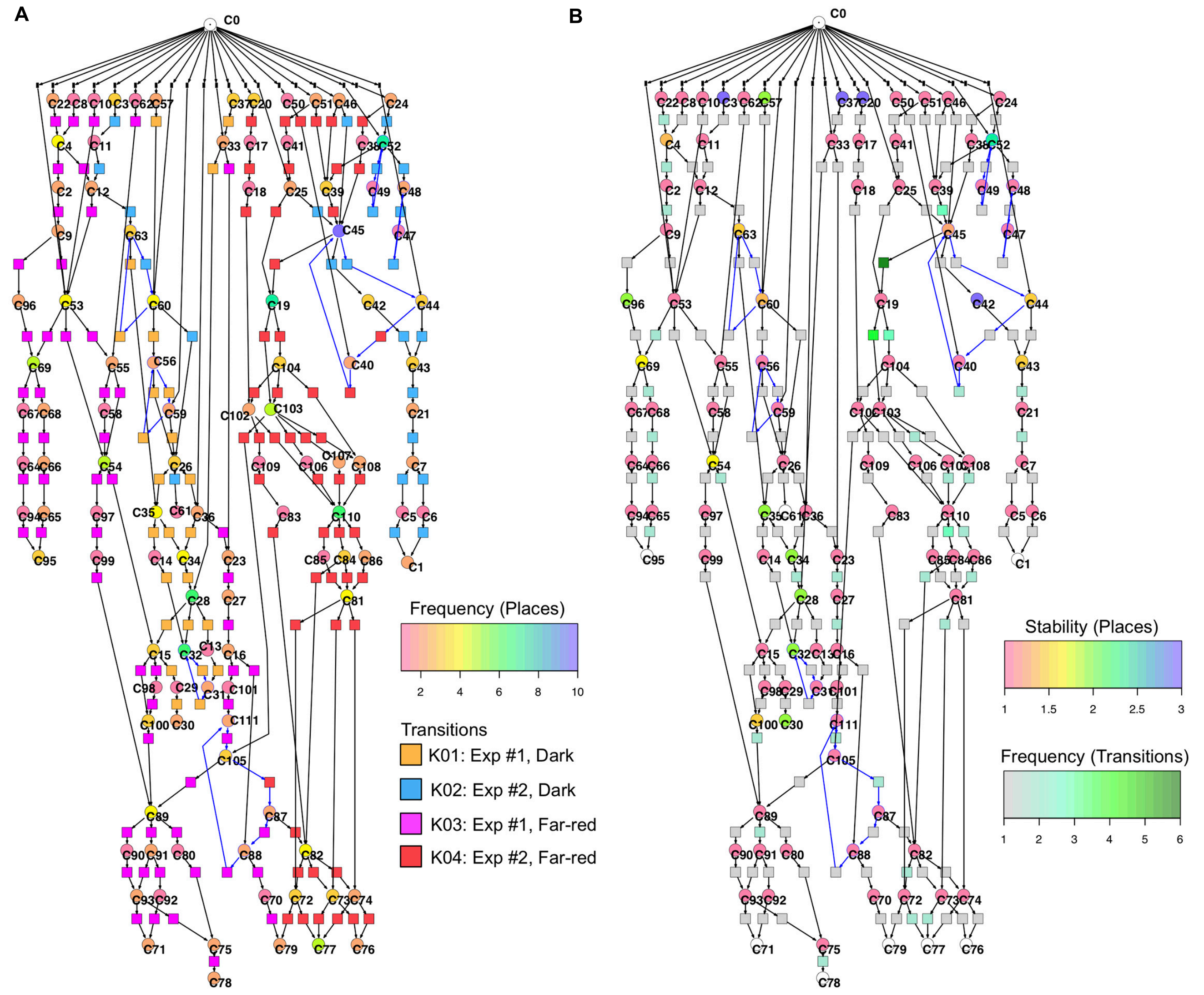


FIGURE 6 | Two copies of the same Petri net, automatically constructed from the single cell trajectories of gene expression as displayed in **Table 2**. Places and transitions in the two nets were colored according to different criteria. Arcs as part of T-invariants are highlighted in blue. **(A)** As indicated by the rainbow color key, places are colored according to the relative frequency with which respective gene expression states occurred in the data set. Transitions are colored according to whether corresponding transits occurred in experiments #1 or #2, and whether cells were far-red stimulated or un-stimulated (dark controls), respectively, as indicated by the panel lower right. **(B)** Places are color coded according to the relative stability of the states of gene expression they represent. This relative stability indicates how long a cell on average resided in a certain state. The color of transitions indicates, in absolute numbers, how frequent a corresponding transit occurred in the data set.

logarithmically, normalized to its concentration at the start of the experiment (t = 0 h) as a function of time for any single cell trajectory. In this kind of plot, the time course of the mRNA of each gene starts at the same point, while the slope of the curve indicates the x-fold change in mRNA abundance over time. Plotting subsets of genes suggests that trajectories through different regions of the Petri net of **Figure 6** indeed emerge from qualitatively different expression kinetics, and that genes are also differently regulated relative to each other when different trajectories are compared. In the example shown in **Figure 9**, *pldA* is early up-regulated in quite a number of trajectories, followed by *pwiA* and finally by *ligA* and *rgsA* that appear strongly correlated at least in some of the plots. Qualitatively different patterns of regulation relative to each other are also evident for the three phospholipase D-encoding genes (**Supplementary Figure 7**). The *pldA* gene is up-regulated while *pldB* and *pldC* are down-regulated. In some of the trajectories, the initial change in the concentration of the *pldA* and *pldC* mRNAs is inverse as compared to the overall time course. A more comprehensive representation with more genes displayed makes similarities and differences between trajectories even more obvious (**Supplementary Figure 8**). Here, we observe a phenomenon, which is also seen in **Table 2**, namely that cells remain in a certain state for some time. This occurs

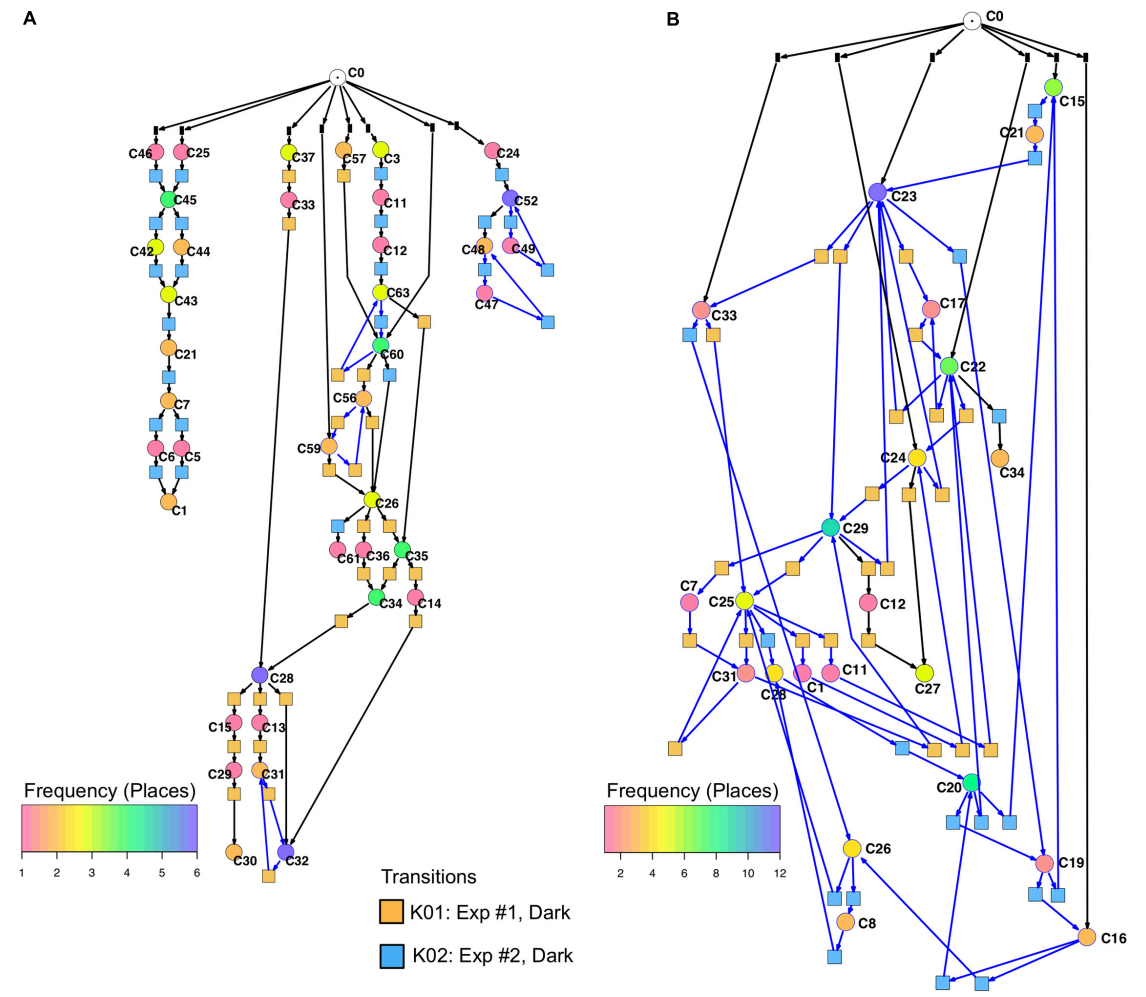


FIGURE 7 | Petri nets constructed for different sets of genes from the gene expression trajectories of un-stimulated cells. **(A)** Trajectories considering the complete set of 35 genes yield three disconnected Petri nets containing a low number of T-invariants (as indicated by arcs highlighted in blue). **(B)** When only those genes are considered that are up- or down-regulated in response to far-red stimulation (omitting the *pcnA*-group of genes), the trajectories of un-stimulated cells give one coherent Petri net which is almost covered with T-invariants. Color coding of places indicates the relative frequencies of states of gene expression in un-stimulated cells. Differences in gene expression patterns of un-stimulated cells between the cells of experiment #1 and #2 are indicated by the appearance of experiment-specific transits and accordingly differently colored transitions.

predominantly in the unstimulated cells but is also seen in some of the light stimulated cells, e.g., in those that proceed to state C95 (**Figures 6A,B**). Cells seemingly are trapped in a meta-stable state (e.g., C96, C69, C100, etc.; **Figure 6B**) for some time until the developmental program proceeds. We presumably will need more data to see whether this is an artifact which occurs by the discretization of gene expression through clustering. Conversely, discretization might help to identify tipping points for the differential regulation of gene expression as the plots in **Supplementary Figure 8** suggest.

DISCUSSION

We have analyzed the gene expression dynamics in response to a differentiation-inducing stimulus pulse in true time by repeatedly taking samples of large, multinucleate plasmodial cells. Control experiments have demonstrated that the gene expression patterns in samples simultaneously retrieved from different sites of a large plasmodial cell did not deviate within the range of the technical accuracy of the measurements. This again confirms that the plasmodial cytoplasm, at least at the level of macroscopic sampling, can be considered as a homogeneous reaction volume.

The injuries caused by multiple sampling of a plasmodial cell heal spontaneously and cutting the plasmodial mass neither induces nor prevents sporulation (Starostzik and Marwan, 1994, 1995b, 1998; Rätzel and Marwan, 2015). This is also confirmed for the large plasmodia used in this study through the dark controls that did not sporulate (**Table 2** and **Supplementary Figure 1**), while the far-red light induced plasmodia sporulated. The changes in gene expression patterns observed in the dark controls seem to be spontaneous and not caused by repeated

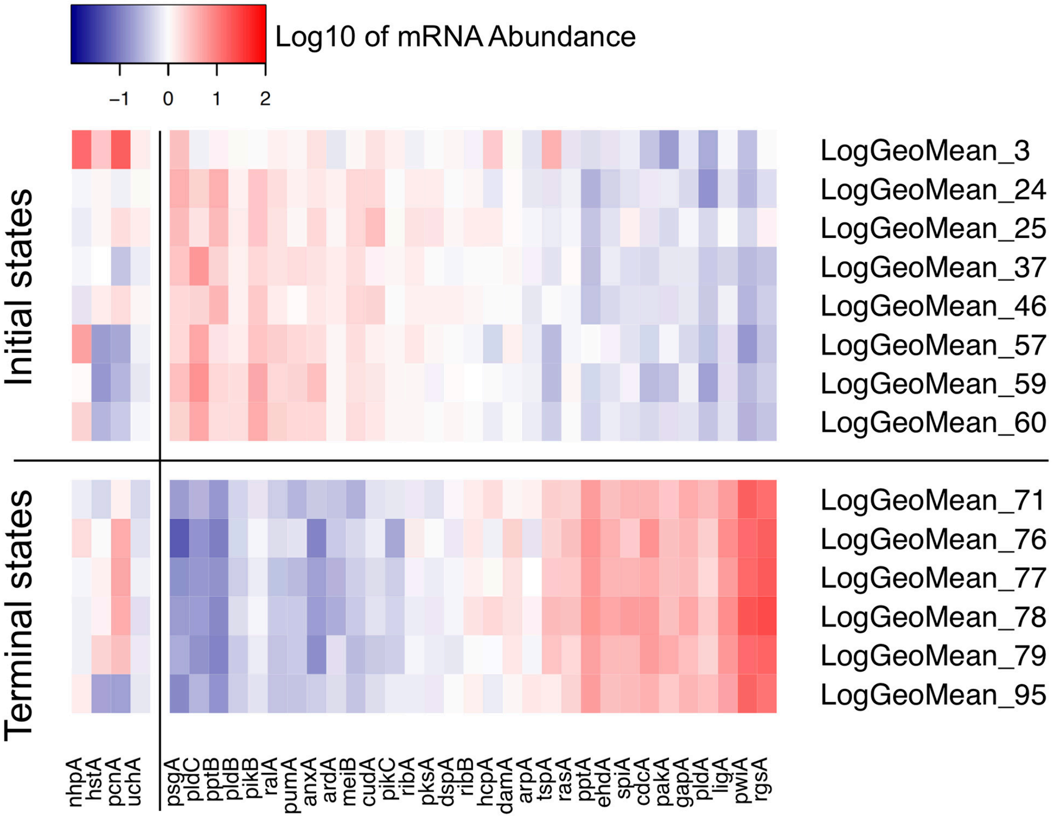


FIGURE 8 | Heat map visualizing the variability of initial and terminal states of gene expression. The initial states displayed in this panel are the start points of trajectories of un-stimulated cells as listed in **Table 2**. The terminal states are endpoints of trajectories of far-red stimulated cells corresponding to terminal places of the Petri net of **Figure 6**. Color-coded expression values correspond to the logarithm to the base 10 of the geometric mean of all expression values of a respective cluster with its ID number indicated on the right side of the panel. For clarity, expression of the *pcnA*-group of genes is displayed as a separate block.

sampling for the following reasons. First, plasmodia are in different states of gene expression already at the start of the experiments (**Figure 8** and **Table 2**), i.e., at the moment before the first sample is taken, so sampling cannot be the reason for this heterogeneity. Second, the trajectories of unstimulated cells developed in different directions although the sampling procedure was the same in all of these cells (**Figures 3A,B**), indicating that the shift in gene expression was not directed with respect to the start of the experiment. Third, the Petri net considering the differential regulation of the subset of light-regulated genes, the expression of which changed to some extent even in unstimulated cells, is almost completely covered with T-invariants (**Figure 7B**), indicating no directed change in the expression of this set of light-regulated genes, neither with respect to the start of the experiment nor with respect to the initial states in which the cells resided before the first sample was taken. We cannot exclude that certain genes might be differentially regulated in response to injury, e.g., genes involved in membrane biosynthesis, as leaks in the membrane readily heal. However, there seems to be no systematic effect on the differential expression of the genes analyzed in the present study.

In contrast to a typical mammalian cell, which has a relatively small cytoplasmic volume, while many genes are present in two copies only, the plasmodial cell contains many millions of nuclei. These nuclei are suspended in a large cytoplasmic volume which continually mixes by the vigorous shuttle streaming. The T-invariants in the Petri nets of **Figure 7B** and **Supplementary Figure 6** indicating the up- and down-regulation of genes were due to changes in the mRNA concentration that occurred at distantly located sites of the plasmodium at the same time. Hence, it seems unlikely that these changes are stochastic gene expression noise. Instead, the T-invariants, presumably do reflect the (non-linear) dynamics of the system. Non-linear, switch-like behavior, bifurcations, multistability, or oscillations all certainly do have fundamental biological and functional implications, and the T-invariant analysis of true single cell time series can help to identify them.

Gene expression states of the cells were defined by hierarchical clustering and discretized by assigning each gene expression pattern to a Simprof significant cluster (Clarke et al., 2008; Rätzel et al., 2020). Trajectories of subsequent discrete states were then assembled into a state machine implemented as a Petri net. In the Petri net, each gene expression state is represented by a place and each transit between two states is represented by a transition.

The Petri net, as it has been defined in this and previous studies depends on the data pre-processing by clustering of the data

(Clarke et al., 2008; Werthmann and Marwan, 2017; Rätzel et al., 2020). It models gene expression trajectories as Markov chains (Gagniuc, 2017), which assumes that each subsequent state only depends on the current state of a cell and not on its previous states, i.e., it does not depend on the individual history of a cell. This assumption is commonly made by computing pseudo-time series from snapshots of individual mammalian cells (Bendall et al., 2014; Shin et al., 2015; Haghverdi et al., 2016; Marr et al., 2016; Street et al., 2018; Weinreb et al., 2018; Chen et al., 2019; Setty et al., 2019). It follows the principle of parsimony in making not more assumptions than necessary and giving the simplest possible explanation for an observed phenomenon. Practically this means that any path which a token can take through the Petri net, by stochastic firing of the transitions, translates into a feasible trajectory of an individual cell. Hence, firing of a transition does only depend on the marking of the pre-place of this transition and not on the identity of any upstream places from which the token originally came. Defining the cell's state of gene expression by measuring more genes might well diversify places and hence change the structure of the Petri net. This has been demonstrated by constructing nets from subsets of genes. In the examples provided, the structure of the Petri net changed and the number of T-invariants increased drastically upon reduction of the number of considered genes (**Table 4**).

If the structure of the Petri net does depend on the set of genes analyzed, what is its actual value? The actual value is that it reveals the behavior of states defined by sets or subsets of genes. Limiting the analysis to the chosen subset of up- and down-regulated genes, as we have done here, revealed extensive on- and off-switching of the genes in unstimulated cells that are differentially regulated in response to a differentiation-inducing stimulus. This became immediately obvious through structural analysis of the Petri net by determining the number of minimal T-invariants. Displaying the net in Sugiyama representation revealed another phenomenon with respect to this subset of genes. The light stimulus caused directed development toward a small number of terminal states reducing the overall number of alternative states in which the cells resided. This suggests that a cellular attractor is formed in response to the stimulus causing the commitment to differentiation.

Coloring the transitions of the Petri net according to the frequency by which transits occurred, allows identification and visualization of main paths, i.e., paths which the system preferably took. Coloring places according to the relative stability of the states they represent indicated metastable states that were not necessarily identical to highly connected places. Places having many pre-transitions (many incoming arcs) represent states, the system is likely to assume, like a corrie in the metaphor of the Waddington landscape, through which the system will pass. Places having many post-transitions (many out-going arcs) represent branching points from which the system has multiple options to proceed.

Our analysis has confirmed former observations (Rätzel et al., 2020), now at considerably larger resolution in time, that unstimulated cells spontaneously and reversibly change their expression pattern. These changes involved the expression of genes that are differentially regulated in response to a differentiation-inducing stimulus. Spontaneous switching of gene expression patterns is at least one reason why stimulated cells started their way to commitment and differentiation from quite different states, indicating substantial heterogeneity in the population of cells. In other words, cells can start differentiation while being in various different states. The differentiation-inducing stimulus then collects or focusses these cells onto a narrow set of states like an attractor of a dynamic system would do. This phenomenon is graphically revealed by the funnel- or cone-like appearance of the Petri net in the Sugiyama layout (**Supplementary Figure 6**).

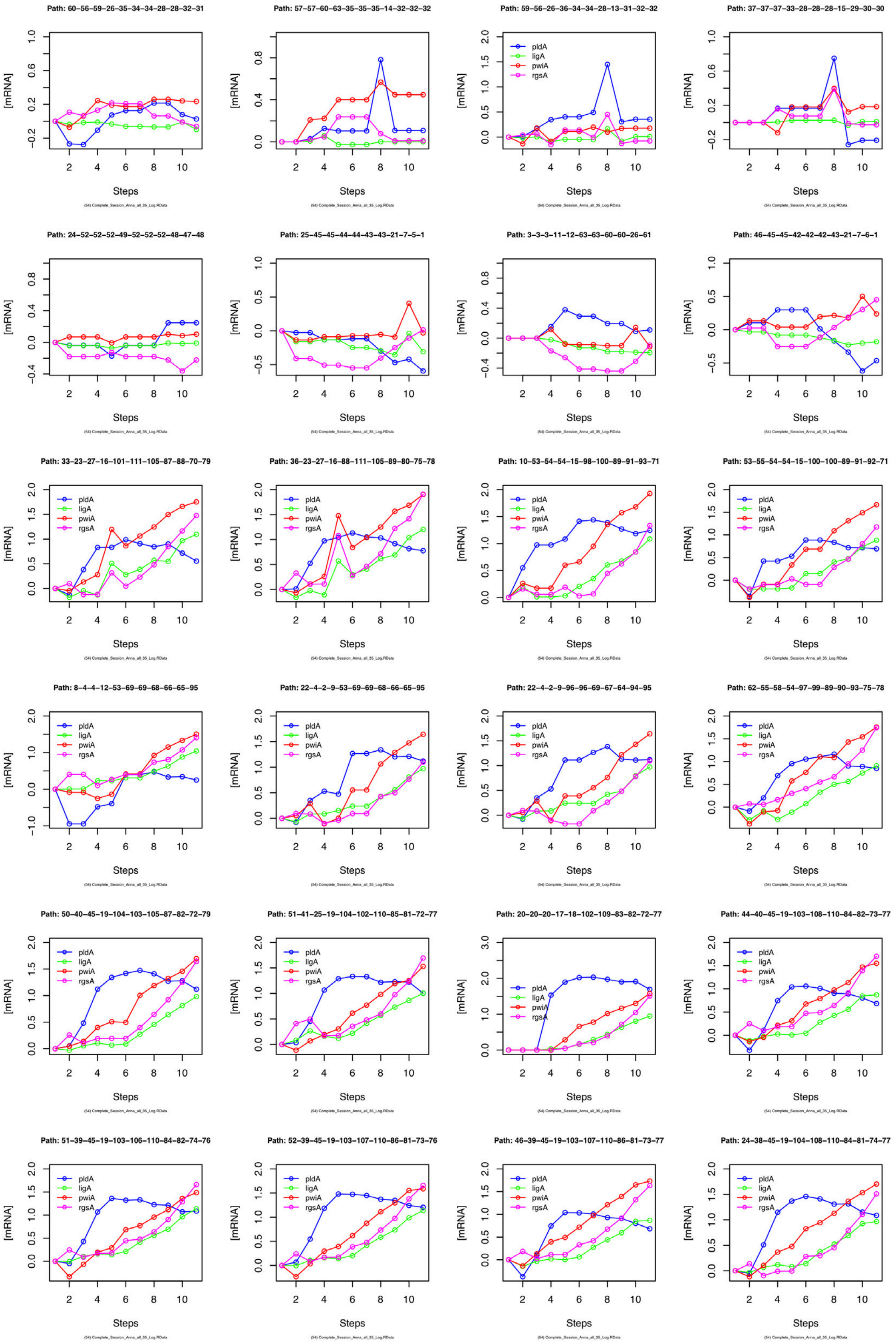
The response of a cell to a differentiation-inducing stimulus seems to depend on the cell's current internal state. **Figure 6A** revealed distinct main branches (visible through transitions of different color) for cells from experiment #1 as compared to experiment #2, suggesting that the response of the cell in terms of its developmental pathway did indeed depend on the initial physiological or gene expression state in which the cell resided while receiving the stimulus. The cells proceeded to slightly different terminal states that however might belong to the same cellular attractor.

One might be tempted to suspect a certain structure in the list of subsequently recorded trajectories (**Table 2**). Changes in the initial state of the plasmodia until the time of stimulus application might have occurred as the experiment proceeded. Similarities in subsequently recorded trajectories may be by chance and we cannot draw any final conclusion because the number of analyzed plasmodia is by far too low. With more plasmodia analyzed and more genes measured, we might discover that the individual history of a cell indeed matters, meaning that the Markov assumption is wrong. Even if this should be the case, the Petri net representation would still be valid, however with the firing probability of certain transitions depending on which path the token came from. Technically, this dependency could be implemented in the form of a colored Petri net. In this context it is trivial and at the same time

TABLE 4 | Number and relative frequency of places, transitions, and T-invariants of Petri nets constructed for different subsets of genes (see legend to **Supplementary Figure 4**) from the data of cells listed in **Table 2**.

Gene set	Down	Up	Up and down	All 35	*pcnA* etc.
Genes	10	10	20	35	4
Places	35	44	69	111	24
Transitions	105	114	144	159	59
P/Gene	3.5	4.4	3.5	3.2	6.0
T/Gene	10.5	11.4	7.2	4.5	14.8
Time points	11	11	11	11	11
Cells	24	24	24	24	24
P/(Genes × tps × cells)	0.013	0.017	0.013	0.012	0.023
T/(Genes × tps × cells)	0.040	0.043	0.027	0.017	0.056
T/P	3.00	2.59	2.09	1.43	2.46
T-Inv	4,413	1,371	1,063	7	732
T-Inv/Gene	441	137	53	0.20	183

P, places; T, transitions; tps, time points; T-Inv, T-invariants.



Path: 60–56–59–26–35–34–34–28–28–32–31

Path: 57–57–60–63–35–35–35–14–32–32–32

Path: 59–56–26–36–34–34–28–13–31–32–32

Path: 37–37–37–33–28–28–28–15–29–30–30

Path: 24–52–52–52–49–52–52–52–48–47–48

Path: 25–45–45–44–44–43–43–21–7–5–1

Path: 3–3–3–11–12–63–63–60–60–26–61

Path: 46–45–45–42–42–42–43–21–7–6–1

Path: 33–23–27–16–101–111–105–87–88–70–79

Path: 36–23–27–16–88–111–105–89–80–75–78

Path: 10–53–54–54–15–98–100–89–91–93–71

Path: 53–55–54–54–15–100–100–89–91–92–71

Path: 8–4–4–12–53–69–69–68–66–65–95

Path: 22–4–2–9–53–69–69–68–66–65–95

Path: 22–4–2–9–96–96–69–67–64–94–95

Path: 62–55–58–54–97–99–89–90–93–75–78

Path: 50–40–45–19–104–103–105–87–82–72–79

Path: 51–41–25–19–104–102–110–85–81–72–77

Path: 20–20–20–17–18–102–109–83–82–72–77

Path: 44–40–45–19–103–108–110–84–82–73–77

Path: 51–39–45–19–103–106–110–84–82–74–76

Path: 52–39–45–19–103–107–110–86–81–73–76

Path: 46–39–45–19–103–107–110–86–81–73–77

Path: 24–38–45–19–104–108–110–84–81–74–77

(54) Complete_Session_Anna_all_35_Log.RData

FIGURE 9 | Gene expression kinetics as derived from single cell trajectories. Each panel represents the trajectory of one single cell as characterized by subsequent states of expression of the same, arbitrarily chosen set of genes. For each gene, the logarithm to the base 10 of the geometric mean of the expression values of each cluster was plotted against time. Plotting the geometric mean instead of individual expression values results in discretization of the data while comparing different trajectories.

important to note that the identity of a place (i.e., state) is always defined by the set of genes that have been measured. Measuring more genes might split any place into more places or even into a separate Petri net with an increased overall number of places. This holds not only for states defined by gene expression but also for cellular states defined by the covalent modification of proteins, *etc*.

We have previously argued that the Petri net depicts aspects of the topology of the Waddington landscape (Waddington, 1957; Huang et al., 2009) with respect to and limited to the set of observed genes (or measured molecular entities) (Werthmann and Marwan, 2017; Rätzel et al., 2020). Then, the token in the Petri net corresponds to the marble rolling down the Waddington landscape as developmental processes unfold. Each Petri net place represents, albeit implicitly, a significantly distinct gene expression state. Despite this implicit representation, the temporal information on each gene for each cell trajectory is available and we have used this information to reveal the temporal hierarchy of differentially regulated genes. The Petri net representation disentangles accordingly the complex gene expression response and identifies alternative regulatory programs or routes. Using this information, the underlying regulatory network can be inferred by applying appropriate algorithms (Marwan et al., 2008; Durzinsky et al., 2011, 2013). This is a possible next step to go.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article are provided as part of the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

AP and SP performed single cell time-series experiments and gene expression analyses and evaluated their results. MHa supervised the experimental work and developed the sample preparation method together with SP. MHe essentially contributed the analysis of the graph-theoretical properties of Petri nets and wrote a corresponding section of the manuscript. WM conceived and supervised the study, performed computational analyses including the automated generation of the Petri nets and wrote the manuscript. All authors read and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.612256/full#supplementary-material

REFERENCES

Alexopoulos, C. J., and Mims, C. W. (1979). *Introductory Mycology*, 3rd Edn. New York,NY: John Wiley Sons.

Aselmeyer, F. (2019). *Analyse des Genexpressionsmusters an unterschiedlichen Stellen von Plasmodien von Physarum polycephalum (II).*Bachelor Thesis Magdeburg: Otto von Guericke Universität.

Bendall, S. C., Davis, K. L., Amir, E.-A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., et al. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157, 714–725. doi: 10.1016/j.cell.2014.04.005

Bornholdt, S., and Kauffman, S. (2019). Ensembles, dynamics, and cell types: Revisiting the statistical mechanics perspective on cellular regulation. *J. Theor. Biol.* 467, 15–22. doi: 10.1016/j.jtbi.2019.01.036

Cannoodt, R., Saelens, W., and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* 46, 2496–2506. doi: 10.1002/eji.201646347

Chen, H., Albergante, L., Hsu, J. Y., Lareau, C. A., Lo Bosco, G., Guan, J., et al. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat. Comm.* 10:1903.

Clarke, K., Somerfield, P., and Gorley, R. (2008). Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *J. Exp. Mar. Biol. Ecol.* 366, 56–69. doi: 10.1016/j.jembe.2008.07.009

Dove, W. F., Dee, J., Hatano, S., Haugli, F. B., and Wohlfarth-Bottermann, K.-E. (eds) (1986). *The Molecular Biology of Physarum polycephalum*. New York,NY: Plenum Press.

Driesch, J. (2019). *Analyse des Genexpressionsmusters an unterschiedlichen Stellen von Plasmodien von Physarum polycephalum (I).*Bachelor Thesis, Magdeburg: Otto von Guericke Universität.

Durzinsky, M., Marwan, W., and Wagler, A. (2013). Reconstruction of extended Petri nets from time-series data by using logical control functions. *J. Math. Biol.* 66, 203–223. doi: 10.1007/s00285-012-0511-3

Durzinsky, M., Wagler, A., and Marwan, W. (2011). Reconstruction of extended Petri nets from time series data and its application to signal transduction and to gene regulatory networks. *BMC Syst. Biol.* 5:113. doi: 10.1186/1752-0509-5-113

Ferrell, J. E Jr. (2012). Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* 22, R458–R466.

Gagniuc, P. A. (2017). *Markov Chains: From Theory to Implementation and Experimentation*. New York, NY: John Wiley & Sons.

Gaublomme Jellert, T., Yosef, N., Lee, Y., Gertner Rona, S., Yang, Li, V., et al. (2015). Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* 163, 1400–1412. doi: 10.1016/j.cell.2015.11.009

Glöckner, G., and Marwan, W. (2017). Transcriptome reprogramming during developmental switching in *Physarum polycephalum* involves extensive remodelling of intracellular signaling networks. *Sci. Rep.* 7:12304.

Golderer, G., Werner, E. R., Leitner, S., Gröbner, P., and Werner-Felmayer, G. (2001). Nitric oxide synthase is induced in sporulation of *Physarum polycephalum*. *Genes Devel.* 15, 1299–1309. doi: 10.1101/gad.890501

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–328. doi: 10.2307/2333639

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. *Nature* 462, 587–594. doi: 10.1038/nature08533

Guttes, E., and Guttes, S. (1961). Synchronous mitosis in starved plasmodia of the myxomycete *Physarum polycephalum*. *Feder. Proc.* 20:419.

Guttes, E., and Guttes, S. (1964). Mitotic synchrony in the plasmodia of *Physarum polycephalum* and mitotic synchronisation by coalescence of microplasmodia. *Meth. Cell Physiol.* 1, 43–54. doi: 10.1016/s0091-679x(08)62085-3

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* 13:845. doi: 10.1038/nmeth.3971

Hayashi, E., Aoyama, N., Wu, Y., Chi, H. C., Boyer, S. K., and Still, D. W. (2007). *Multiplexed, quantitative gene expression analysis for lettuce seed germination on GenomeLabTM GeXP genetic analysis system. Beckman Coulter Application Information A-10295A*. Available online at https://ls.beckmancoulter.co.jp/files/appli_note/A_10295A.pdf (accessed December 16, 2020).

Heiner, M. (2009). "Understanding Network Behavior by Structured Representations of Transition Invariants," in *Algorithmic Bioprocesses. Natural Computing Series*, eds A. Condon, D. Harel, J. Kok, A. Salomaa, and E. Winfree, (Berlin: Springer), 367–389. doi: 10.1007/978-3-540-88869-7_19

Heiner, M., Rohr, C., and Schwarick, M. (2013). "MARCIE – Model Checking and Reachability Analysis Done Efficiently," in *Application and Theory of Petri Nets* *and Concurrency. PETRI NETS 2013. Lecture Notes in Computer Science*, Vol. 7927, eds J. M. Colom, and J. Desel, (Berlin: Springer).

Hoffmann, X.-K., Tesmer, J., Souquet, M., and Marwan, W. (2012). Futile attempts to differentiate provide molecular evidence for individual differences within a population of cells during cellular reprogramming. *FEMS Microbiol. Lett.* 329, 78–86. doi: 10.1111/j.1574-6968.2012.02506.x

Huang, S., Ernberg, I., and Kauffman, S. (2009). Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* 20, 869–876. doi: 10.1016/j.semcdb.2009.07.003

Il Joo, J., Zhou, J. X., Huang, S., and Cho, K.-H. (2018). Determining relative dynamic stability of cell states using boolean network model. *Sci. Rep.* 8:12077.

Junker, J. P., and van Oudenaarden, A. (2014). Every cell is special: Genome-wide studies add a new dimension to single-cell biology. *Cell* 157, 8–11. doi: 10.1016/j.cell.2014.02.010

Kafri, R., Levy, J., Ginzberg, M. B., Oh, S., Lahav, G., and Kirschner, M. W. (2013). Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* 494, 1–4. doi: 10.1201/b14602-2

Lamparter, T., and Marwan, W. (2001). Spectroscopic detection of a phytochrome-like photoreceptor in the myxomycete *Physarum polycephalum* and the kinetic mechanism for the photocontrol of sporulation by P_{fr}. *Photochem. Photobiol.* 73, 697–702. doi: 10.1562/0031-8655(2001)073<0697:sdoapl>2.0.co;2

Lautenbach, K. (1973). Exact liveness conditions of a Petri net class. GMD Report 82, Bonn (in German).

Macaulay, I. C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., et al. (2016). Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in hematopoietic cells. *Cell Rep.* 14, 966–977. doi: 10.1016/j.celrep.2015.12.082

Marco, E., Karp, R. L., Guo, G., Robson, P., Hart, A. H., Trippa, L., et al. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci. U S A* 111, E5643–E5650.

Marquardt, P., Werthmann, B., Raetzel, V., Haas, M., and Marwan, W. (2017). Quantifying 35 transcripts in a single tube: Model-based calibration of the GeXP RT-PCR assay. *bioRxiv*. doi: 10.1101/159723

Marr, C., Zhou, J. X., and Huang, S. (2016). Single-cell gene expression profiling and cell state dynamics: collecting data, correlating data points and connecting the dots. *Curr. Opin. Biotechnol.* 39, 207–214. doi: 10.1016/j.copbio.2016.04.015

Marwan, W., Wagler, A., and Weismantel, R. (2008). A mathematical approach to solve the network reconstruction problem. *Math. Meth. Oper. Res.* 67, 117–132. doi: 10.1007/s00186-007-0178-5

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 33, 269–276. doi: 10.1038/nbt.3154

Moris, N., Pina, C., and Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* 7, 693–703. doi: 10.1038/nrg.2016.98

Paul, F., Arkin, Ya, Giladi, A., Jaitin Diego, A., Kenigsberg, E., et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* 163, 1663–1677. doi: 10.1016/j.cell.2015.11.013

Plass, M., Solana, J., Wolf, F. A., Ayoub, S., Misios, A., Glažar, P., et al. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* 360:eaaq1723. doi: 10.1126/science.aaq1723

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rätzel, V. (2015). *Dynamische Fließgleichgewichte und ihre Übergänge in Reaktionsnetzwerken: Experimenteller Nachweis der Quasi-potential-Landschaft der zellulären Reprogrammierung*. Magdeburg: Otto von Guericke University.

Rätzel, V., and Marwan, W. (2015). Gene expression kinetics in individual plasmodial cells reveal alternative programs of differential regulation during commitment and differentiation. *Dev. Growth Differ.* 57, 408–420. doi: 10.1111/dgd.12220

Rätzel, V., Werthmann, B., Haas, M., Strube, J., and Marwan, W. (2020). Disentangling a complex response in cell reprogramming and probing the Waddington landscape by automatic construction of Petri nets. *BioSystems* 189:104092. doi: 10.1016/j.biosystems.2019.104092

Rohr, C., Marwan, W., and Heiner, M. (2010). Snoopy–a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics* 26, 974–975. doi: 10.1093/bioinformatics/btq050

Rusch, H. P., Sachsenmaier, W., Behrens, K., and Gruter, V. (1966). Synchronization of mitosis by the fusion of the plasmodia of *Physarum polycephalum*. *J. Cell Biol.* 31, 204–209. doi: 10.1083/jcb.31.1.204

Sackmann, A., Heiner, M., and Koch, I. (2006). Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinform.* 7:482. doi: 10.1186/1471-2105-7-482

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* 37, 547–554. doi: 10.1038/s41587-019-0071-9

Schaap, P., Barrantes, I., Minx, P., Sasaki, N., Anderson, R. W., Bénard, M., et al. (2016). The *Physarum polycephalum* genome reveals extensive use of prokaryotic two-component and metazoan-type tyrosine kinase signaling. *Genome Biol. Evol.* 8, 109–125. doi: 10.1093/gbe/evv237

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* 37, 451–460. doi: 10.1038/s41587-019-0068-4

Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., et al. (2015). Single-cell RNA-seq with Waterfall reveals molecular cascades underlying adult neurogenesis. *Stem Cell* 17, 360–372. doi: 10.1016/j.stem.2015.07.013

Spiller, D. G., Wood, C. D., Rand, D. A., and White, M. R. H. (2010). Measurement of single-cell dynamics. *Nature* 465, 736–745.

Starostzik, C., and Marwan, W. (1994). Time-resolved detection of three intracellular signals controlling photomorphogenesis in *Physarum polycephalum*. *J. Bacteriol.* 176, 5541–5543. doi: 10.1128/jb.176.17.5541-5543.1994

Starostzik, C., and Marwan, W. (1995a). Functional mapping of the branched signal transduction pathway that controls sporulation in *Physarum polycephalum*. *Photochem. Photobiol.* 62, 930–933. doi: 10.1111/j.1751-1097.1995.tb09158.x

Starostzik, C., and Marwan, W. (1995b). A photoreceptor with characteristics of phytochrome triggers sporulation in the true slime mould *Physarum polycephalum*. *FEBS Lett.* 370, 146–148. doi: 10.1016/0014-5793(95)00820-y

Starostzik, C., and Marwan, W. (1998). Kinetic analysis of a signal transduction pathway by time-resolved somatic complementation of mutants. *J. Exp. Biol.* 201, 1991–1999.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., et al. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19:477. doi: 10.1186/s12864-018-4772-0

Sugiyama, K., Tagawa, S., and Toda, M. (1981). Methods for visual understanding of hierarchical system structures. *IEEE Transac. Syst. Man Cyber.* 11, 109–125. doi: 10.1109/tsmc.1981.4308636

Waddington, C. H. (1957). *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin.

Walter, P., Hoffmann, X.-K., Ebeling, B., Haas, M., and Marwan, W. (2013). Switch-like reprogramming of gene expression after fusion of multinucleate plasmodial cells of two *Physarum polycephalum* sporulation mutants. *Biochem. Biophys. Res. Comm.* 435, 88–93. doi: 10.1016/j.bbrc.2013.04.043

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., et al. (2016). *gplots: Various R Programming Tools for Plotting Data. R package version 3.0*.

Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M., and Klein, A. M. (2018). Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci.* 115, E2467–E2476.

Werthmann, B., and Marwan, W. (2017). Developmental switching in *Physarum polycephalum*: Petri net analysis of single cell trajectories of gene expression indicates responsiveness and genetic plasticity of the Waddington quasipotential landscape. *J. Phys. D Appl. Phys.* 50:464003. doi: 10.1088/1361-6463/aa8e2b

Whitaker, D., and Christman, M. (2014). *clustsig: Significant Cluster Analysis. R package version 1.1*.

Zhou, J. X., and Huang, S. (2011). Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet.* 27, 55–62.

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

frontiers in Physiology

ORIGINAL RESEARCH
published: 28 January 2021
doi: 10.3389/fphys.2020.596263

Spindle Architectural Features Must Be Considered Along With Cell Size to Explain the Timing of Mitotic Checkpoint Silencing

Mathew Bloomfield, Jing Chen* and Daniela Cimini*

Department of Biological Sciences and Fralin Life Sciences Institute, Virginia Tech, Blacksburg, VA, United States

Mitosis proceeds through a defined series of events that is largely conserved, but the amount of time needed for their completion can vary in different cells and organisms. In many systems, mitotic duration depends on the time required to satisfy and silence the spindle assembly checkpoint (SAC), also known as the mitotic checkpoint. Because SAC silencing involves trafficking SAC molecules among kinetochores, spindle, and cytoplasm, the size and geometry of the spindle relative to cell volume are expected to affect mitotic duration by influencing the timing of SAC silencing. However, the relationship between SAC silencing, cell size, and spindle dimensions is unclear. To investigate this issue, we used four DLD-1 tetraploid (4N) clones characterized by small or large nuclear and cell size. We found that the small 4N clones had longer mitotic durations than the parental DLD-1 cells and that this delay was due to differences in their metaphase duration. Leveraging a previous mathematical model for spatiotemporal regulation of SAC silencing, we show that the difference in metaphase duration, i.e., SAC silencing time, can be explained by the distinct spindle microtubule densities and sizes of the cell, spindle, and spindle poles in the 4N clones. Lastly, we demonstrate that manipulating spindle geometry can alter mitotic and metaphase duration, consistent with a model prediction. Our results suggest that spindle size does not always scale with cell size in mammalian cells and cell size is not sufficient to explain the differences in metaphase duration. Only when a number of spindle architectural features are considered along with cell size can the kinetics of SAC silencing, and hence mitotic duration, in the different clones be explained.

Keywords: mitosis, cell size, nuclear size, mitotic spindle, mitotic checkpoint, tetraploidy, SAC

Edited by:
Denis Tsygankov,
Georgia Institute of Technology,
United States

Reviewed by:
Nathan Weinstein,
Universidad Nacional Autónoma de
México, Mexico
Victor M. Bolanos-Garcia,
Oxford Brookes University,
United Kingdom

***Correspondence:**
Jing Chen
chenjing@vt.edu
Daniela Cimini
cimini@vt.edu

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: *18 August 2020*
Accepted: *23 December 2020*
Published: *28 January 2021*

Citation:
Bloomfield M, Chen J and Cimini D
(2021) Spindle Architectural Features
Must Be Considered Along With Cell
Size to Explain the Timing of Mitotic
Checkpoint Silencing.
Front. Physiol. 11:596263.
doi: 10.3389/fphys.2020.596263

INTRODUCTION

Mitosis requires the completion of specific events in a timely manner; the precise sequence of events governs progression through defined mitotic stages (Baudoin and Cimini, 2018). Although these mitotic stages are conserved in different cell types and organisms, the time scale for their completion varies, with mitosis typically lasting 10–20 min in *Drosophila* embryos (McCleland et al., 2009), *S. pombe* (Krüger et al., 2019), and *S. cerevisiae* (Brewer et al., 1984; Leitao and Kellogg, 2017), 20–60 min in many mammalian cell lines (Rieder et al., 1994; Meraldi et al., 2004; Arnaoutov et al., 2005; Kuznetsova et al., 2015; Viganó et al., 2018), and 1–2 h in mouse embryos (Sikora-Polaczek et al., 2006). In many systems, mitotic duration critically depends on the spindle assembly checkpoint (SAC), the surveillance mechanism that monitors kinetochore-microtubule attachments and halts mitotic progression until all kinetochores are bound to spindle microtubules (Musacchio, 2011). There are two key SAC-regulated events whose duration can influence mitotic timing. First, the time it takes to satisfy the SAC by establishing kinetochore-microtubule attachments and chromosome biorientation (Gorbsky et al., 1998; Hauf et al., 2003; Mogilner and Craig, 2010; Foley and Kapoor, 2013; Sacristan and Kops, 2015). This process defines the stage of mitosis known as prometaphase, during which SAC signaling remains active. Second, the time it takes for the SAC to be silenced after complete chromosome alignment at the metaphase plate has been achieved (Rieder et al., 1994, 1995; Howell et al., 2000; Shah et al., 2004; Pereira and Maiato, 2012). This time would define the duration of metaphase. During prometaphase, unattached kinetochores serve as a platform to promote a conformational change in the SAC protein Mad2, which is then able to bind other SAC proteins in the cytoplasm and form the mitotic checkpoint complex (MCC), which in turn inhibits the anaphase promoting complex/cyclosome (APC/C), producing the so-called "wait-anaphase" signal (Taylor et al., 2004; Musacchio, 2015). This signal spreads throughout the spindle and cytoplasm (Heasley et al., 2017) and is capable of inducing a mitotic arrest in response to a single unattached kinetochore (Rieder et al., 1994, 1995). Once a kinetochore achieves stable attachment to microtubules, key SAC proteins are stripped from the attached kinetochore and transported poleward along spindle microtubules by the motor protein dynein (Howell et al., 2001), and SAC-activating phosphorylation events are reversed by phosphatases at the kinetochore (Etemad and Kops, 2016; Moura et al., 2017; Gelens et al., 2018; Saurin, 2018). Dynein-mediated transport is important for timely silencing of the SAC signal after all kinetochores achieve stable attachment to the microtubules (Howell et al., 2001; Griffis et al., 2007; Gassmann et al., 2010). SAC silencing leads to activation of the APC/C, degradation of securin and cyclin B, and consequently chromosome segregation and mitotic exit (Clute and Pines, 1999; Hagting et al., 2002). While stable kinetochore-microtubule attachments are sufficient to satisfy the SAC in mammalian cells (Tauchman et al., 2015), it is still unclear how robust and efficient silencing of the SAC is achieved.

Because SAC silencing involves trafficking SAC molecules among kinetochores, spindle, and cytoplasm, cell and spindle sizes are expected to affect mitotic duration by influencing the time required for SAC silencing. Recent studies have started to provide insight on some facets of this issue, but questions still remain. For instance, mitotic spindle size is thought to adjust to cell size (Heald and Gibeaux, 2018). This is evident in the early stages of embryogenesis, when developing cells rapidly divide without growth, producing cells of progressively smaller size with shorter spindles (Loughlin et al., 2010; Good et al., 2013; Hazel et al., 2013; Reber et al., 2013; Wilbur and Heald, 2013; Lacroix et al., 2018). Differences in genome size or ploidy may also affect spindle length indirectly by altering cell size (Mortimer, 1958; Mayer et al., 1992; Edgar and Orr-Weaver, 2001; Gregory, 2001; Gillooly et al., 2015). This has been observed in *Xenopus*, where cell and spindle sizes are larger in the near-tetraploid *X. laevis* compared to its diploid relative *X. tropicalis* (Brown et al., 2007; Loughlin et al., 2011). Furthermore, spindle scaling has been observed in a variety of metazoan species with different cell sizes (Crowder et al., 2015). However, in experimentally-generated tetraploid budding yeast strains, spindle length failed to scale with cell volume (Storchová et al., 2006), showing that we do not yet fully understand how the mitotic spindle scales in response to changes in genome and cell size. Similarly, the question of whether changes in cell/spindle size influence SAC silencing is still awaiting a definitive answer. A theoretical model predicts that spindle size scaling is important for robust SAC silencing (Chen and Liu, 2016). However, experimental investigation of this question has often been partial or indirect and has yielded contrasting observations. For instance, cell cycle and mitotic progression were not perturbed in tetraploid yeast cells (Storchová et al., 2006), which may be due to the lack of spindle scaling in those strains. In contrast, tetraploid mammalian cells often progress through mitosis more slowly than their diploid counterparts (Kuznetsova et al., 2015; Paim and FitzHarris, 2019; Cohen-Sharir et al., 2020; Quinton et al., 2020), suggesting a possible delay in SAC silencing. Several recent studies, which explored the link between cell and spindle sizes and SAC function more closely (Gerhold et al., 2015, 2018; Galli and Morgan, 2016; Kyogoku and Kitajima, 2017), also produced conflicting results. Gerhold et al. showed that depleting HIM-10, a component of the NDC80 complex, increased spindle length and metaphase duration in *C. elegans*, and the metaphase delay was only observed in cells with functional Mad2 (Gerhold et al., 2015). Conversely, mouse oocytes with excessively large cytoplasmic and spindle volumes progressed into anaphase more rapidly than smaller oocytes (Kyogoku and Kitajima, 2017). To further complicate the issue, many studies that investigate the relationship between cell size and the SAC are performed under conditions of mitotic spindle disruption (Galli and Morgan, 2016; Gerhold et al., 2018; Vázquez-Diez et al., 2019)—a method used to quantify SAC strength by measuring the time required for mitotic exit after depolymerization of spindle microtubules (Rieder and Maiato, 2004; Khodjakov and Rieder, 2009). This is especially common in studies examining SAC silencing, where Mps1 inhibition is combined with spindle disruption to avoid any effects kinetochore-microtubule attachments may have on the signaling events that prompt anaphase onset (Nijenhuis et al., 2014; Sivakumar et al., 2014; Saurin, 2018; Smith et al., 2019). Therefore, a thorough understanding of whether and how the time required for SAC silencing depends on cell and spindle sizes in the presence of an intact mitotic spindle is lacking. Moreover, previous studies have not investigated the contribution of other spindle architectural features, such as its shape and density, to SAC silencing and mitotic duration.

To study the effects of cell and spindle size on SAC silencing, we generated tetraploid (4N) clones from non-transformed, immortalized retinal pigmented epithelial (RPE-1) and DLD-1 colorectal cancer cells. Not all of the 4N DLD-1 clones displayed the same scaling of cell and spindle size compared to the diploid parental cells, which made them an ideal model to study the relationship between cell/spindle size and mitotic duration. We found that the 4N clones had unique spindle geometries and certain clones exhibited altered mitotic timings, which were due to differences in metaphase duration (i.e., time taken to silence the SAC after all kinetochores are attached). Leveraging a previous mathematical model for spatiotemporal regulation of SAC silencing (Chen and Liu, 2014, 2016), we show that the differences in SAC silencing time, and hence metaphase duration, cannot be explained by cell and spindle size alone. Instead, only when other spindle architectural features, including spindle pole size and microtubule abundance, are considered along with spindle and cell size can the timing of SAC silencing and metaphase duration be explained in all clones. Lastly, in line with a model prediction, we demonstrate that experimentally manipulating spindle geometry can reduce metaphase duration in a clone characterized by a particularly long metaphase.

MATERIALS AND METHODS

Cell Lines and Culture Conditions

DLD-1 (ATCC CCL-221) and hTERT-immortalized RPE-1 cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA). The DLD-1 cells were maintained in RPMI 1640 media with ATCC modification (Thermo Fisher Scientific—Gibco, CA, USA) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific) and 1% antibiotic-antimycotic (Thermo Fisher Scientific), and the RPE-1 cells were maintained in a 1:1 mixture of DMEM/F-12 with HEPES (Thermo Fisher Scientific). All cells were kept in a humidified incubator at 37°C and 5% CO_2 and monitored for possible mycoplasma infection every 2–3 weeks by DNA staining.

Tetraploid RPE-1 and DLD-1 cells were generated by treating diploid cells with 1.5 μg/mL dihydrocytochalasin B (DCB; Sigma Aldrich, St. Louis, MO) for 20 hrs. After treatment, the cells were washed four times with cell culture media and allowed to grow an additional 1–2 days in supplemented media before the isolation of single cells by limiting dilution in 96-well-plates. Only wells containing a single cell were expanded into clonal cell lines and used for further experimentation. Characterization of 4N clones and subsequent experiments were performed at low passages to limit variability due to evolution of the cell population.

For experiments conducted in the presence of blebbistatin (Tocris Bioscience, Bristol, UK), the drug was added to the media at a final concentration of 10 nM for 4 h before fixation and immunostaining or at the time of media replacement prior to live-cell imaging.

Preparation of Chromosome Spreads and Chromosome Counting

Cells were grown to 70–80% confluency in T-25 flasks and 50 ng/ml colcemid (Karyomax–Invitrogen, Waltham, MA) was added to the media for 4–6 h to enrich for mitotic cells. The cells were collected by trypsinization and centrifuged at 1,000 rpm for 5 min. The cell pellet was resuspended in 2 mL PBS and centrifuged for 3 min at 1,000 rpm. Five milliliters of hypotonic solution (0.075 M KCl) was added to the cell pellet and incubated for 18 min at 37°C; then 0.5 mL of freshly prepared fixative (3:1 methanol:glacial acetic acid) was added before centrifugation at 1,000 rpm for 5 min. After removal of the supernatant, the cell pellet was resuspended by adding 5 mL of fixative dropwise and then incubated at room temperature for 15 min before centrifugation at 1,000 rpm for 5 min. Depending on the size, resulting cell pellets were suspended in 0.3–6 mL fixative (added dropwise) and 12 μL of the cell suspension were dropped onto microscope slides to check for optimal density and spreading. Chromosome spreads were left at room temperature overnight and stained the next day with 300 nM DAPI (Invitrogen) for 10 min. Antifade solution (90% glycerol and 0.5% N-propyl gallate) was added to the slides and sealed under a 22 × 50 mm coverslip (Corning Incorporated, Corning, NY) with nail polish. Chromosome spreads were imaged using a Nikon Eclipse Ti inverted microscope (Nikon Instruments Inc., NY, USA) equipped with ProScan automated stage (Prior Scientific, Cambridge, UK), CoolSNAP HQ2 CCD camera (Photometrics, AZ, USA), Lumen200PRO light source (Prior Scientific), and a 60X/1.4 NA Plan-Apochromatic objective. Chromosomes were counted in individual images of chromosome spreads using NIS Elements AR version 4.60 software (Nikon Instruments Inc.) to confirm cell ploidy.

Cell and Nuclear Volume Measurements

For cell and nuclear volume measurement, cells were synchronized in G2 to reduce variation in size due to cell cycle stage. Cells were seeded onto sterilized acid-washed glass coverslips inside 35 mm Petri dishes at low densities to enable cells to adhere without touching neighbors. The next day, 9 μM of the CDK-1 inhibitor RO-3306 (Sigma Aldrich) (Vassilev, 2006) was added to the media for 18 h to synchronize cells in G2. After synchronization, the cells were incubated in serum-free media containing 5 μM CellTracker Green CMFDA Dye (Thermo Fisher Scientific) for at least 30 min to stain the cytoplasm. The serum-free medium was replaced with medium containing FBS for at least 30 min before fixation per the manufacturer's protocol. Cells were fixed in ice-cold methanol for 10 min at −20°C, washed three times (5 min each) with PBS, and counterstained with 300 nM DAPI (Invitrogen) for 5 min. Coverslips were mounted on microscope slides in an antifade solution and sealed with nail polish. Z-stack images spanning the entire height of single cells were acquired at 0.6 μm steps with a swept field confocal system (Prairie Technologies, WI, USA) on a Nikon Eclipse TE2000 inverted microscope equipped with a 60X/1.4 NA Plan-Apochromatic lens, motorized ProScan stage (Prior Scientific), an XCITE 120Q light source (Excelitas Technologies, Waltham, MA, USA; used to view cells before imaging), a CoolSNAP HQ2 CCD camera (Photometrics), an Agilent monolithic laser combiner (MLC400) controlled by a four channel acousto-optic tunable filter, and a multiband pass filter set (illumination at 405, 488, 561, and 640 nm). Cell and nuclear volume measurements were performed in FIJI [ImageJ, NIH (Rasband, 2011)] using a macro for three-dimensional reconstruction. In brief, image smoothing was done using a Gaussian function, a binary image of each Z-stack was created by auto thresholding (v1.17.2, "default" FIJI algorithm) with the stack histogram, any holes were filled, and the "3D Object Counter" was used to quantify 3D objects in each stack and determine cell and nuclear volumes (Bolte and Cordelières, 2006).

Immunofluorescence, Image Acquisition, and Data Analysis

For spindle dimension measurements, prometaphase-metaphase defect analysis, and mitotic stage analysis, cells were fixed in freshly prepared 4% paraformaldehyde in PHEM buffer (60 mM Pipes, 25 mM HEPES, 10 mM EGTA, 2 mM $MgSO_4$, pH 7.0) for 20 min at room temperature. Cells were then washed three times (5 min each) with PBS and permeabilized in PHEM buffer containing 0.5% Triton-X 100 for 10 min at room temperature. After three quick washes with PBS, cells were blocked with 20% boiled goat serum for 1 h at room temperature. For cyclin B immunostaining and α- and γ-tubulin fluorescence intensity measurements, cells were fixed using ice-cold methanol for 10 min at −20°C and washed three times with PBS. For γ-tubulin, cells were blocked for 1 h at room temperature with 10% BGS. For cyclin B and α-tubulin, 20% BGS was used for blocking. Next, the cells were incubated overnight at 4°C with primary antibodies in 5% BGS and PHEM buffer. Cells were then washed four times in PBS-T (PBS with 0.05% Tween 20) before 45 min incubation at room temperature with secondary antibodies in 5% BGS and PHEM buffer. The following primary antibodies and dilutions were used: mouse anti-centrin (Abnova, Zhongli, Taiwan), 1:200; rabbit anti-α-tubulin (Abcam, Cambridge, MA), 1:250; human anti-centromere protein (Antibodies Inc., Davis, CA), 1:100; mouse anti-α-tubulin (DM1A, Sigma Aldrich), 1:500; rabbit anti-γ-tubulin (Abcam), 1:200; mouse anti-cyclin B (BD Biosciences, San Jose, CA), 1:500. The following secondary antibodies and dilutions were used: Alexa 488 goat anti-rabbit (Molecular Probes, Life Technologies, CA, USA), 1:200; Rhodamine Red-X goat anti-human (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA), 1:100; Cy5 goat anti-mouse (Abcam), 1:200; Alexa 488 goat anti-mouse (Molecular Probes), 1:200.

To analyze the mitotic spindle, Z-stack images were acquired at 0.6 μm steps using the Nikon Eclipse TE2000 microscope setup described in the previous section. For spindle width, a maximum projection image was used to determine the distance between the outermost kinetochores in the metaphase plate by drawing and measuring a straight line in FIJI. For spindle length, a straight line was drawn and measured from the most intense centrin signal at each spindle pole. For cells where the spindle poles were not in the same focal plane, spindle length was calculated using the Pythagorean theorem. Spindle height was determined as the distance from the first to the last plane of the Z planes containing kinetochores. Spindle volume was calculated as: π/12 × spindle length × spindle width × spindle height.

For quantification of cells with unaligned chromosomes, the analysis was limited to late prometaphase cells (i.e., cells with a well-defined metaphase plate and only 1-few unaligned chromosomes). For analysis of multipolarity, analysis was limited to cells in which a metaphase plate was discernible. For mitotic stage analysis, the number of mitotic cells in prophase, prometaphase, metaphase, anaphase, and telophase were counted (Baudoin and Cimini, 2018).

For fluorescence intensity analysis, images for all cell lines in each experiment were captured on the same day using the same microscope and imaging settings and were analyzed using FIJI software. For quantification of γ-tubulin intensity, Z planes containing a detectable γ-tubulin signal (as determined by the default auto thresholding function in FIJI) were converted to a single image using a sum intensity projection. 17-pixel (I_a) and 24-pixel (I_b) diameter circles centered around the γ-tubulin signal were used to determine background-subtracted spindle pole fluorescence intensity using the following formula:

$$\text{Spindle Pole Fluorescence Intensity} = \mathrm{I_a} - (\mathrm{I_b} - \mathrm{I_a}).$$

For spindle microtubule measurements, auto thresholding of a maximum projection of *Z* planes spanning the entire spindle was performed to create a binary image. Any holes were filled and the "analyze particles" function was used to define the spindle area, which was saved as region of interest (ROI). Next, this ROI was overlaid onto a sum intensity projection of *Z* planes spanning the entire spindle height to measure the integrated density of all pixels within the spindle. For background correction, two identical ROIs were drawn outside of the spindle area, but within the cell, to measure intracellular background fluorescence. These measurements were averaged (total background fluorescence/total background ROI area), then this value was extrapolated to match spindle area, and subtracted from the α-tubulin spindle intensity using the following formulas:

$$\text{SpindleBackground Intensity} = \text{average background intensity} \times \text{spindle area}$$

$$\text{SpindleMicrotubule Intensity} = \text{Total } \alpha - \text{tubulin Intensity} - \text{Spindle Background Intensity}$$

Live-Cell Phase Contrast Microscopy

Cells were grown on glass bottom dishes with No. 1.5 glass (MatTek Corporation, Ashland, MA). Immediately prior to imaging, cell medium was replaced with L-15 media (Gibco) supplemented with 4.5 g/L glucose. All live-cell experiments were performed on the same Nikon Eclipse Ti inverted microscope (Nikon Instruments Inc.) described earlier ("Preparation of chromosome spreads and chromosome counting") with a temperature and humidity-controlled incubator (Tokai Hit, Japan). For live-cell phase contrast videos, images were acquired every 3 min using a 20X/0.3 NA A Plan corrected phase contrast objective for at least 6 h. Videos were analyzed with NIS Elements to determine mitotic duration, or the time from cell rounding to anaphase onset. In some cells, especially the larger 4N groups, nuclear envelope breakdown (NEBD) could also be observed and typically coincided with the beginning of cell rounding. For analysis of prometaphase and metaphase duration, only cells with visible metaphase plates were used for analysis. Prometaphase-metaphase transition was recorded as the time a metaphase plate could be seen at the spindle equator.

Statistical Analysis

Statistical analysis was performed using GraphPad Prism (version 8.3.1) software. For prometaphase-metaphase defect and mitotic stage analysis, statistical significance was determined using a two-sided Fisher's exact test. For all other experiments, a Student's *t*-test was used to test for statistical significance. For experiments in which individual data points are reported, results from individual experiments were compared to each other for reproducibility and confirm that no statistical differences were observed between experiments.

Mathematical Model for Spatiotemporal Regulation of SAC Silencing

The model for SAC silencing consists of a core model for the spatial dynamics of SAC proteins, and coupled biochemical circuits for SAC signaling and silencing (Chen and Liu, 2014). The core model predicts the concentration of SAC proteins at the spindle pole as increasing number of kinetochores establish attachment to the spindle and was built upon the following key assumptions motivated by experimental observations:

(i) High phosphorylation level at unattached kinetochores vs. low phosphorylation level at attached ones (Liu et al., 2009; Maresca and Salmon, 2009; Welburn et al., 2010) dictates the distinct dynamics and fates of SAC proteins at the two types of kinetochores. Because of the known interdependency between kinetochore attachment state, phosphorylation level of kinetochore proteins, and dynein-mediated transport (Whyte et al., 2008; Matson and Stukenberg, 2014), the model assumes that SAC proteins released from the attached and unattached kinetochores adopt distinct states. The SAC proteins released from an attached kinetochore assume a transport-active state, forming complexes with active dynein and undergoing poleward streaming, while those released from an unattached kinetochore assume a transport-inactive state with diffusion only. Moreover, because kinetochore phosphorylation level promotes recruitment of SAC proteins (Ditchfield et al., 2003), the model assumes a significant decrease in the recruitment rate of SAC proteins as a kinetochore turns to the attached state.

(ii) The transport-active SAC proteins move along microtubules toward the spindle pole and they may unbind before reaching the pole due to dynein's limited processivity (King and Schroer, 2000; Reck-Peterson et al., 2006). While unbound from microtubules, these proteins diffuse in the cytoplasm and have a chance to either rebind the microtubules and keep moving toward the spindle pole or bind a kinetochore. Once these proteins reach the spindle pole, they are partially sequestered by the spindle pole through binding/unbinding dynamics.

The core model was formulated as a system of compartmentalized diffusion-advection-reaction partial differential equations (PDE) given in the **Supplementary Materials** (Chen and Liu, 2014, 2015, 2016). The core model was further expanded to an extended model, by coupling the spatial dynamics with a biochemical circuit that describes the regulatory interactions between key molecules of the SAC and a generic toggle switch circuit (Tyson et al., 2003) that is triggered by a threshold SAC protein concentration at the spindle pole and induces SAC silencing. The PDEs for the extended model are also provided in the **Supplementary Materials**.

Size and Microtubule Parameters Used in the Model

Because cell/spindle sizes and mitotic timing must be measured with several different experimental setups and could not all be obtained from the same cell, we could only use the model to make predictions based on the average sizes in each clone. To avoid large deviation caused by outlier data points, we used the median values of the measured sizes and fluorescence intensities. The measured values and derived model parameters are summarized in **Table 1**.

RESULTS

Since tetraploidy may have disparate effects on cell and spindle size that disrupt mitotic fidelity (Storchová et al., 2006), we reasoned it could be used to study the relationship between cell size, mitotic spindle geometry, and SAC function. Therefore, we induced cytokinesis failure with dihydrocytochalasin B in RPE-1 (diploid, non-transformed) and DLD-1 cells (pseudodiploid colorectal cancer cells) and expanded clonal cell lines from single tetraploid (4N) cells. We were able to isolate one 4N RPE-1 clone and several 4N DLD-1 clones. We next characterized these 4N clones to determine nuclear and cell volumes in G2-synchronized cells. In RPE-1 cells, cell and nuclear volume scaled with DNA content (**Supplementary Figures 1A,B**, **Table 2**). For DLD-1 cells, however, we found that, although chromosome number doubled in each DLD-1 4N clone (**Figure 1A**, **Table 2**), cell and nuclear volume did not scale uniformly (**Figures 1B–D**, **Table 2**). Two of the DLD-1 4N clones were characterized by small nuclear and cell size (S clones) and two were characterized by large sizes (L clones). Specifically, the nuclear volume of L1 and L2 increased by 96 and 95%, respectively, compared to parental DLD1 cells, consistent with the increase in DNA content; however, S1 and S2 only increased by 37 and 38%, respectively (**Figures 1B,C**, **Table 2**). Similarly, the cell volume of L1 and L2 increased by 86 and 98%, respectively, while S1 and S2 increased by 28 and 31%, respectively (**Figures 1B,C**, **Table 2**). Notably, cell and nuclear volume scaled proportionally with each other, such that the nucleus-to-cytoplasm (N/C) ratio of the respective parental cells was largely maintained in each 4N clone (**Table 2**). These findings show that tetraploidy has a variable effect on cell and nuclear volumes in DLD-1 cells and that doubling of the genome does not always result in doubling of nuclear size. Finally, we should note that in all 4N clones the

TABLE 1 | Median values of measured quantities in cell clones and parameters used in the model.

	Quantity	DLD-1	S1	S2	L1	L2	RPE-1 (2N)	RPE-1 (4N)
Measured	Cell Volume (μm^3)	1,847	2,227	2,383	3,153	3,081	2,399	5,595
	Spindle length (μm)	6.83	7.27	10.33	9.86	12.20	11.82	12.23
	Spindle width (μm)	8.85	9.97	10.74	10.94	11.44	8.69	11.17
	Spindle height (μm)	6.60	9.00	9.60	7.80	6.60	6.00	7.80
	γ-tubulin intensity (a.u.)	864.81	1,595.0	888.87	1,373.8	1,309.1	934.87	999.97
	α-tubulin intensity (a.u.)	893.27	1,841.0	1,030.6	2,201.1	1,080.6	959.92	971.45
Model	Cell diameter (μm): $D = \sqrt[3]{measured\ cell\ volume \times 6 \div \pi}$	15.22	16.20	16.57	18.19	18.05	16.61	22.03
	Spindle length (μm): $L = measured\ spindle\ length$	6.83	7.27	10.33	9.86	12.20	11.82	12.23
	Spindle width/height (μm): $W = \sqrt{\frac{measured\ spindle\ width \times}{measured\ spindle\ height}}$	7.64	9.47	10.15	9.24	8.69	7.22	9.33
	Spindle pole diameter (μm), *d* [see **Supplementary Material**, Eq. (S1)]	1.11	1.50	1.12	1.39	1.36	1.57	1.62
	Number of kinetochore microtubules: N_{MT1} (see **Supplementary Material**, Eq. (S2))	1,600	3,193	916	2,885	960	1,032	776

TABLE 2 | Characterization of parental DLD-1 and RPE-1 cells and derived tetraploid clones.

Characteristic	DLD-1	S1	S2	L1	L2	RPE-1 (2N)	RPE-1 (4N)
Modal Chromosome No.	46	91	92	88	91	46	88
Nuclear Volume* (μm^3)	847 ± 39	1,156 ± 41	1,170 ± 60	1,656 ± 107	1,648 ± 117	649 ± 27	1,465 ± 47
Cell Volume* (μm^3)	1,787 ± 80	2,288 ± 97	2,345 ± 123	3,326 ± 233	3,541 ± 244	2,403 ± 95	5,731 ± 270
N/C Ratio	0.48	0.51	0.50	0.51	0.47	0.27	0.27

**Nuclear and cell volumes are reported as mean ± SEM.*

majority of cells had normal centrosome numbers, consistent with a recent report that extra centrosomes are rapidly lost in 4N cells via asymmetric centrosome clustering during bipolar cell divisions and selective advantage of cells inheriting a single centrosome (Baudoin et al., 2020). During such bipolar divisions with asymmetric centrosome clustering, the chromosomes are partitioned in a bipolar manner and cytokinesis occurs at the midzone. Therefore, this process is unlikely to explain the size differences we observe in our clones, but it is consistent with the emergence of 4N cell populations in which the vast majority of cells assemble bipolar spindles with normal centrosome numbers (Ganem et al., 2009; Godinho et al., 2014; Kuznetsova et al., 2015; Potapova et al., 2016; Viganó et al., 2018; Baudoin et al., 2020).

Tetraploid Clones Display Distinct Mitotic Spindle and Cell Geometries

Spindle length increases linearly with cell diameter in many metazoan species (Crowder et al., 2015), but in 4N budding yeast, spindle length did not increase compared to smaller diploid and haploid counterparts (Storchová et al., 2006). To determine if the mitotic spindle scales with cell size in our 4N clones, we measured spindle length, width, and height in metaphase cells with immunostained microtubules (α-tubulin), kinetochores (centromere antigen), and centrioles (centrin) (**Figure 2A**). Spindle width, measured as the distance between the outermost kinetochores at the metaphase plate, increased by roughly 20% in all the RPE-1 and DLD-1 4N clones compared to their respective parental cells (**Figure 2B**, **Supplementary Figure 1C**), suggesting that this width may be optimal for tetraploid cells regardless of their size and may be necessary to accommodate the extra chromosomes. Spindle height, which was also measured using kinetochores at the metaphase plate, increased by 25% in the RPE-1 4N clone compared to the RPE-1 diploid cells (**Supplementary Figure 1D**). In the 4N DLD-1 clones, however, spindle height did not significantly change in L2 and increased by only 17% in L1 compared to the parental cells (**Figure 2C**). Instead, in both small DLD-1 4N clones (S1 and S2), spindle height increased by 33–34% compared to the DLD-1 parental cells (**Figure 2C**). Finally, we measured spindle length as the distance between the spindle poles (defined by centriole staining) and found that it did not increase in the RPE-1 4N clone compared to its diploid ancestor (**Supplementary Figure 1E**), despite the considerable increase in cell volume (**Supplementary Figure 1B**). Compared to the diploid DLD-1 cells, spindle length increased by 37 and 71%,

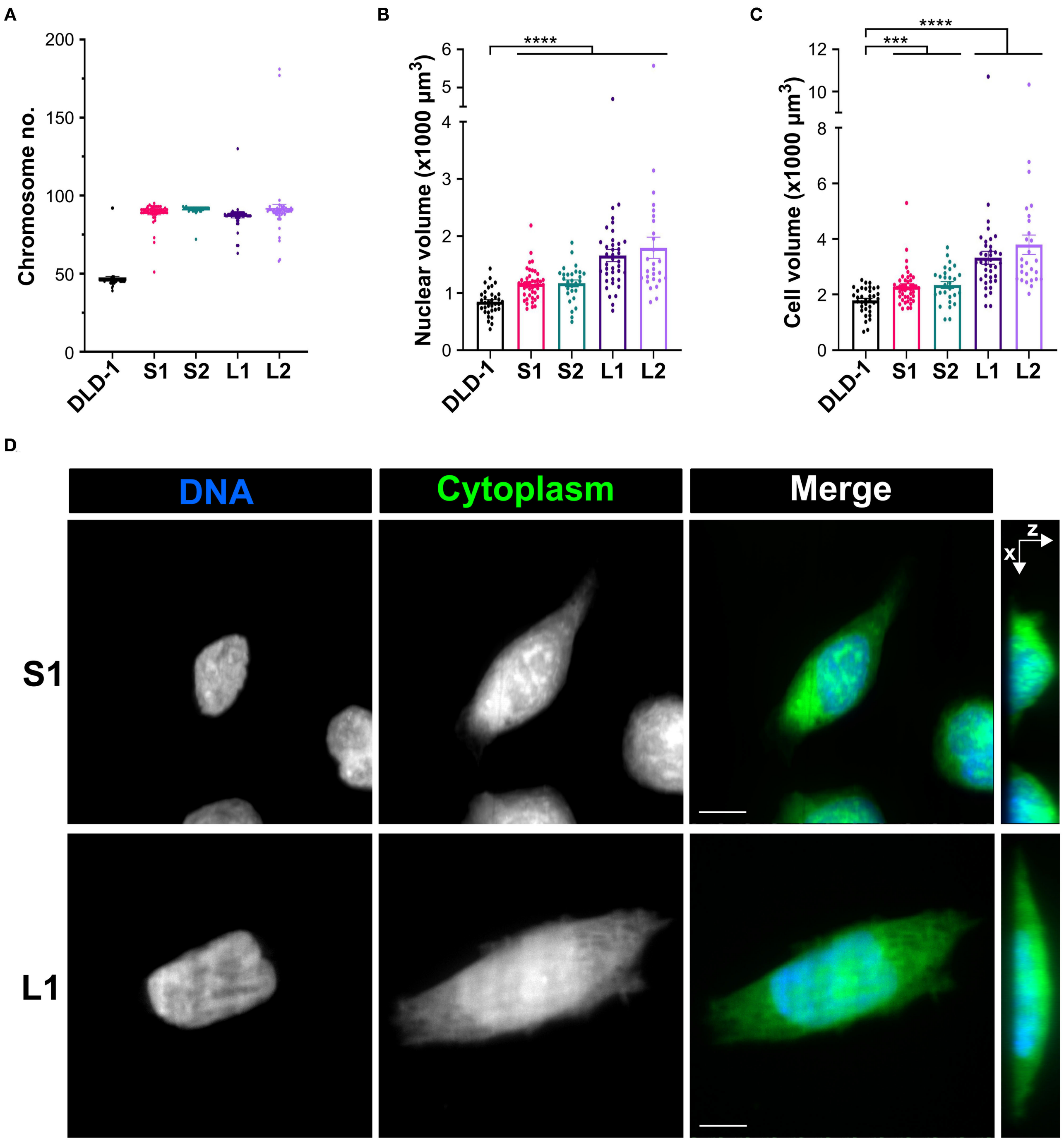


FIGURE 1 | Tetraploidization has a variable effect on cell and nuclear volume in DLD-1 cells. **(A)** Analysis of chromosome number to confirm ploidy in DLD-1 cells and derived tetraploid clones ($n = 50$ for each). **(B)** Nuclear and **(C)** cell volume measurements in parental DLD-1 cells and 4N DLD-1 clones synchronized in G2. Data are reported as mean ± SEM with individual data points from three independent experiments in which a total of 27–44 cells ($n = 35$, 44, 28, 39, and 27, respectively) were analyzed. ***$p < 0.001$, ****$p < 0.0001$, when compared to the parental DLD-1 cells by Student's *t*-test. **(D)** Examples of S1 (top) and L1 (bottom) cells used for nuclear (blue) and cell (green) volume analysis. Images are maximum intensity projections from *Z*-stacks along the *X*–*Y* coordinates (left three columns) or along the *X*–*Z* coordinates (right column). All scale bars, 10 µm.

respectively, in the two large clones L1 and L2 (**Figure 2D**). In the two small clones, S1 and S2, spindle length increased by 7.7 and 42%, respectively (**Figure 2D**), indicating that spindle length can vary widely between tetraploid cells of similar sizes. These findings show that tetraploidy can have mixed effects on spindle length and height, producing distinct spindle geometries. The RPE-1 4N clone increased spindle width and height, but not spindle length, relative to the parental cells. Compared to diploid DLD-1, L1 and L2 primarily expanded spindle length, whereas S2 expanded both spindle height and spindle length, and S1 expanded spindle height, but not spindle length, resulting in a spindle with a "compressed" appearance (**Supplementary Videos 1–10**).

To determine if spindle volume correlates with cell volume, we calculated the volume of the entire mitotic spindle using the spindle dimensions of each 4N clone. In the RPE-1 4N clone, spindle volume did not scale with cell volume, increasing by only 60% relative to the RPE-1 cells (**Supplementary Figure 1F**). For the DLD-1 clones, spindle volume increased by 91 and 109% in L1 and L2, respectively (**Figure 2E**), compared to the parental DLD-1 cells, similar to the increases in cell volume (**Figures 1B,C**). In the small 4N DLD-1 clones, however, spindle volume scaling exceeded the changes in cell volume, with S1 and S2 increasing by 61 and 125%, respectively (**Figure 2E**), compared to parental DLD-1 cells. Moreover, despite being similar in size (**Figures 1B–D**), S1 and S2 had the smallest and largest spindle, respectively, of all the 4N DLD-1 clones (**Figure 2E**), suggesting that cell size is not the only determinant of spindle volume. Overall, these results show that in response to tetraploidy, spindle volume did not fully scale with cell size in the tetraploid RPE-1 cells, whereas in DLD-1 cells, although the 4N clones adjusted their spindle geometries differently, spindle volume mirrored or surpassed the increases in cell volume.

Since the large 4N DLD-1 clones mostly expanded spindle length and width rather than height, we asked whether there were also differences in the sphericity of metaphase cells in the 4N DLD-1 clones. Therefore, we measured the cell diameter (along the axis corresponding to spindle length) of metaphase cells. Cell diameter increased in L1 and L2 by about 25% compared to the parental DLD-1 cells; however, cell diameter did not change in S2 and decreased by 20% in S1 compared to the DLD-1 cells (**Supplementary Figure 2A**). To assess sphericity, we used the G2 cell volume measurements (from **Figure 1C**) to calculate the expected diameter assuming the cells formed perfect spheres during mitosis and compared this to our observed metaphase cell diameters (**Supplementary Figure 2B**). This showed that the observed and expected cell diameters were more similar in the small 4N DLD-1 clones compared to the large 4N DLD-1 clones or the DLD-1 cells (**Supplementary Figure 2B**), indicating that S1 and S2 cells are more spherical while L1 and L2 cells are more ellipsoid during mitosis. In particular, the observed and expected cell diameters in S1 were nearly identical (**Supplementary Figure 2B**), suggesting these cells form almost perfect spheres when dividing. Overall, these findings show a clear correlation between the size of tetraploid DLD-1 cells and the degree of mitotic rounding.

Small Tetraploid Cells Spend More Time in Mitosis Due to a Delay in Metaphase

Previous studies have shown that, in some experimental systems, tetraploidy can lengthen mitosis, but it is unclear if cell and spindle scaling contribute to this delay (Kuznetsova et al., 2015; Viganó et al., 2018; Paim and FitzHarris, 2019; Cohen-Sharir et al., 2020; Quinton et al., 2020). To investigate whether the differences in cell size and spindle geometry among the 4N clones affect their progression through mitosis, we measured the timing from cell rounding to anaphase onset (referred to as mitotic duration henceforth) using live-cell phase contrast microscopy (**Figure 3A** and **Supplementary Videos 11–15**). Mitotic duration was similar in the 2N and 4N RPE-1 cells (**Supplementary Figure 3A**). Similarly, mitotic duration did not significantly change in L1 or L2 (43.5 and 45.0 min, respectively), but it was significantly longer in the small 4N clones S1 (59.8 min) and S2 (47.2 min) compared to DLD-1 cells (41.1 min; **Figure 3B**).

Since the small 4N DLD-1 clones showed a delay in mitosis, we next sought to identify the underlying cause of the delay. Chromosome misalignment can delay mitosis (Potapova and Gorbsky, 2017), so we performed fixed-cell analysis (**Figure 4A**, top row) to quantify unaligned chromosome(s) in late prometaphase cells (i.e., cells with a well-defined metaphase plate and only 1 or a few unaligned chromosomes). Among all the 4N clones, L1, L2, and the 4N RPE-1, but not S1 and S2, displayed significantly higher frequencies of cells with unaligned chromosomes compared to the parental cells (**Figure 4B** and **Supplementary Figure 3D**), indicating that the mitotic delay in S1 and S2 is not caused by a delay in chromosome alignment. Transient spindle multipolarity (**Figure 4A**, bottom row) can also lengthen mitosis (Silkworth et al., 2009). However, L1, L2, and 4N RPE-1, but not S1 and S2, displayed an increase in multipolar prometaphase/metaphase cells compared to the parental cells (**Figure 4C** and **Supplementary Figure 3E**), indicating that transient multipolarity cannot explain the mitotic delay in the small 4N clones.

Overall, the above findings suggest the mitotic delay in small 4N DLD-1 clones may occur during metaphase. To explore this possibility, we used fixed-cell mitotic stage analysis to quantify the percentage of mitotic cells in metaphase (**Figure 4D**). We found similar percentages of metaphase cells in 4N RPE-1, L1, and L2 cells compared to their parental cells (**Figure 4D** and **Supplementary Figure 3F**). Conversely, the fraction of metaphase cells was significantly higher in S1 and S2 compared to the parental cells, suggesting that the delay in mitotic duration in the small 4N DLD-1 clones could be due to prolonged time in metaphase.

To further assess whether metaphase duration was prolonged in S1 and S2, we measured the time from cell rounding to metaphase (prometaphase) and from metaphase to anaphase onset (metaphase) by live-cell phase contrast microscopy. Consistent with the similar overall mitotic timing, 2N and 4N RPE-1 cells had comparable prometaphase durations of nearly 15 min (**Supplementary Figure 3B**). The parental DLD-1 cells had an average prometaphase duration of 21.2 min, which was similar to the prometaphase duration for L1 (21.4 min; **Figure 3C**). A modest increase in prometaphase duration was observed for all other 4N DLD-1 clones, where prometaphase duration was 23.4 (S2), 24.2 (S1), and 25.4 (L2) min (**Figure 3C**). However, these differences are not sufficient to explain the overall increase in mitotic duration

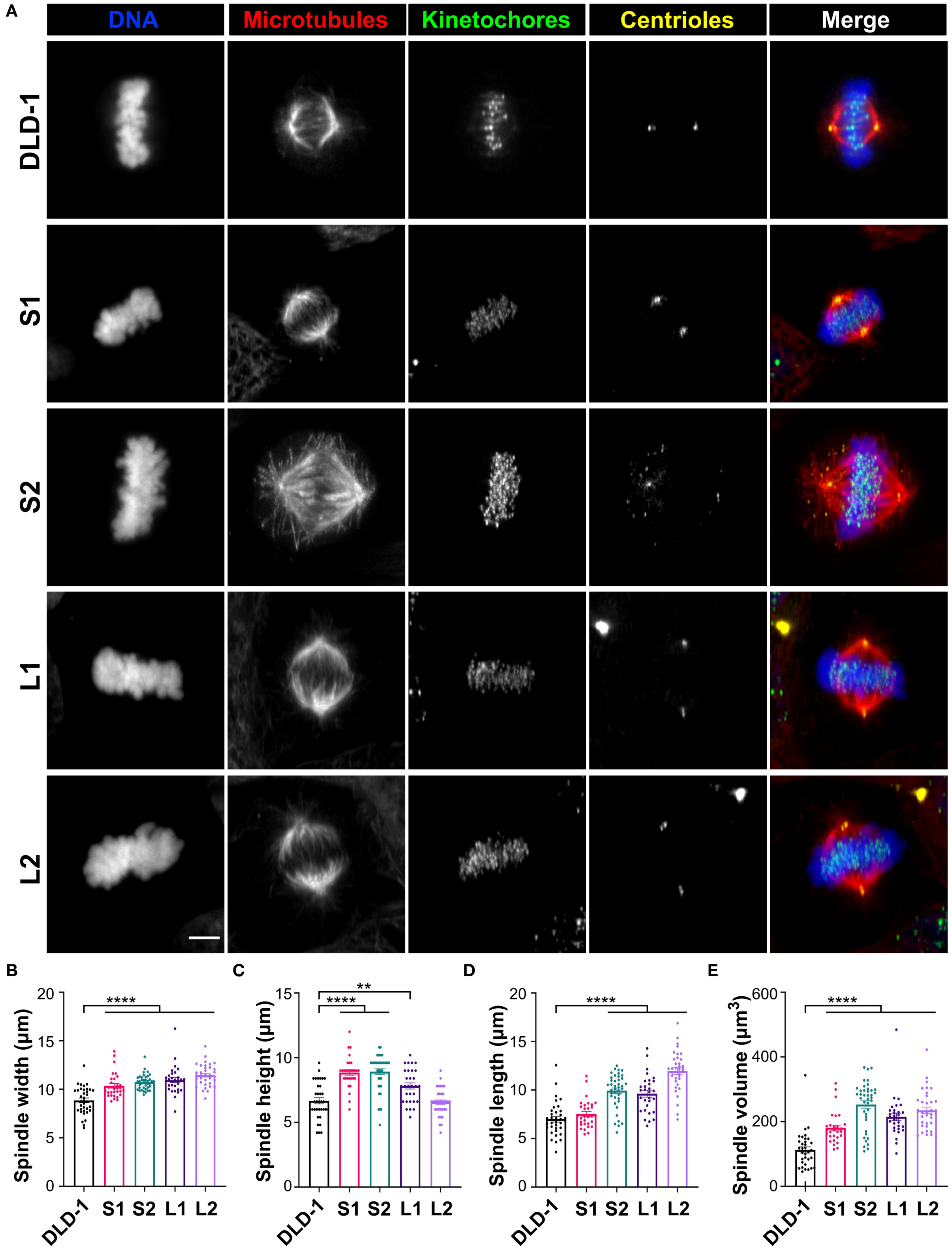


FIGURE 2 | Tetraploid DLD-1 clones have distinct mitotic spindle geometries. **(A)** Examples of mitotic spindles from parental DLD-1 cells and cells from 4N DLD-1 clones S1, S2, L1, and L2 (top-to-bottom) in metaphase. Scale bar, 5 µm. Measurements of mitotic spindle width **(B)**, height **(C)**, length **(D)**, and volume **(E)** reported as mean ± SEM with individual data points from three independent experiments in which a total of 30–40 cells (*n* = 35, 30, 40, 30, and 35, respectively). $^{**}p < 0.01$, $^{****}p < 0.0001$, when compared to the parental DLD-1 cells by Student's *t*-test.

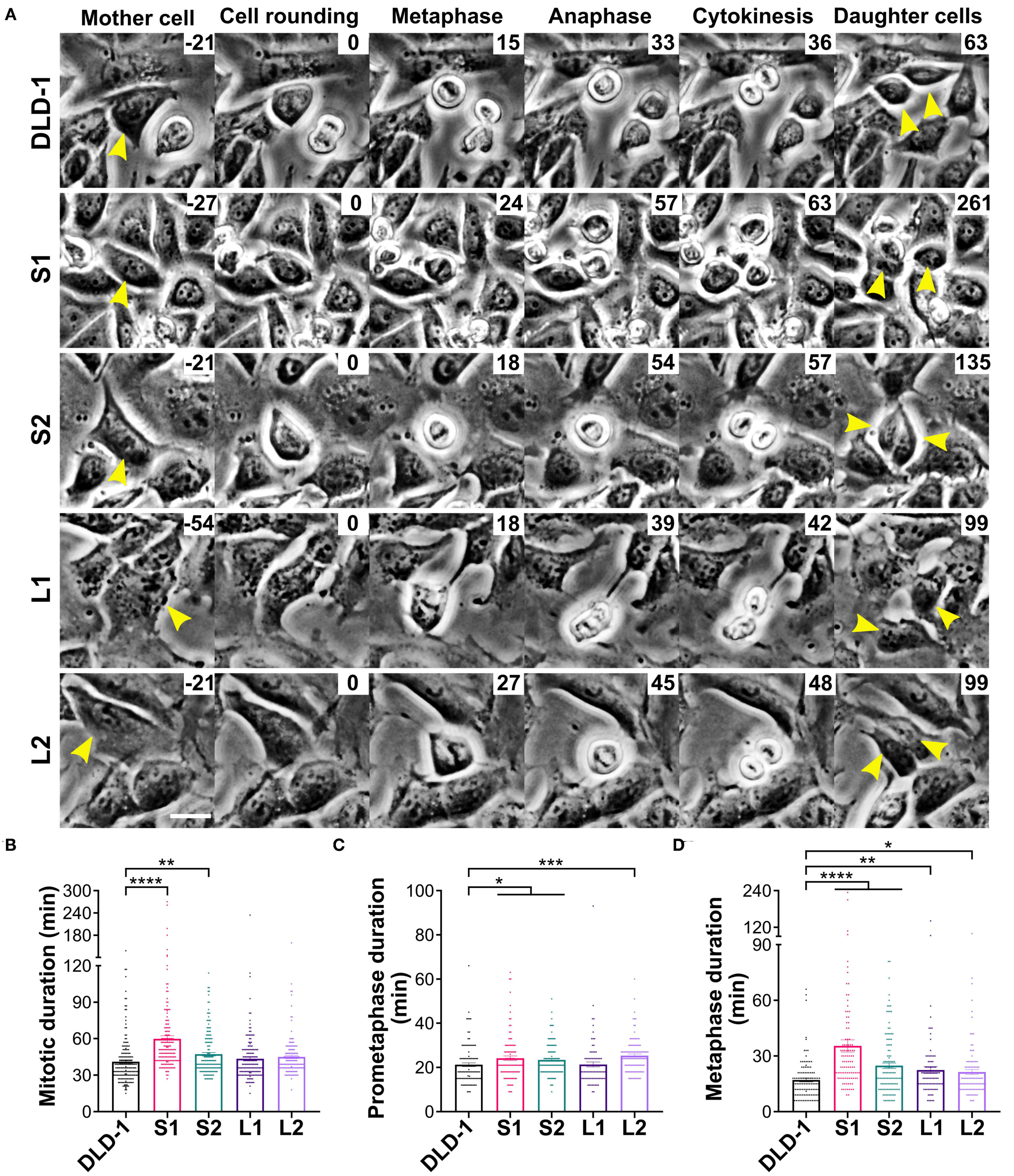


FIGURE 3 | Small 4N DLD-1 cells spend more time in mitosis due to a delay in metaphase. **(A)** Examples of parental DLD-1 cells and cells from 4N DLD-1 clones S1, S2, L1, and L2 (top-to-bottom) progressing through mitosis. Elapsed time is shown in minutes (top-right corner), with 0 min corresponding to cell rounding. Yellow arrowheads point to the dividing cell (Mother cell column) and to the two daughter cells (right column). Scale bar, 25 µm. Quantification of **(B)** mitotic duration ($n =$ 179, 167, 161, 161, and 150 cells, respectively), **(C)** prometaphase duration ($n =$ 111, 101, 118, 90, and 110 cells, respectively), and **(D)** metaphase duration ($n =$ 111, 101, 121, 91, and 101 cells, respectively). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$, when compared to the parental DLD-1 cells by Student's *t*-test. Graphs in panels **(B–D)** represent data collected from three independent experiments reported as mean ± SEM with individual data points.

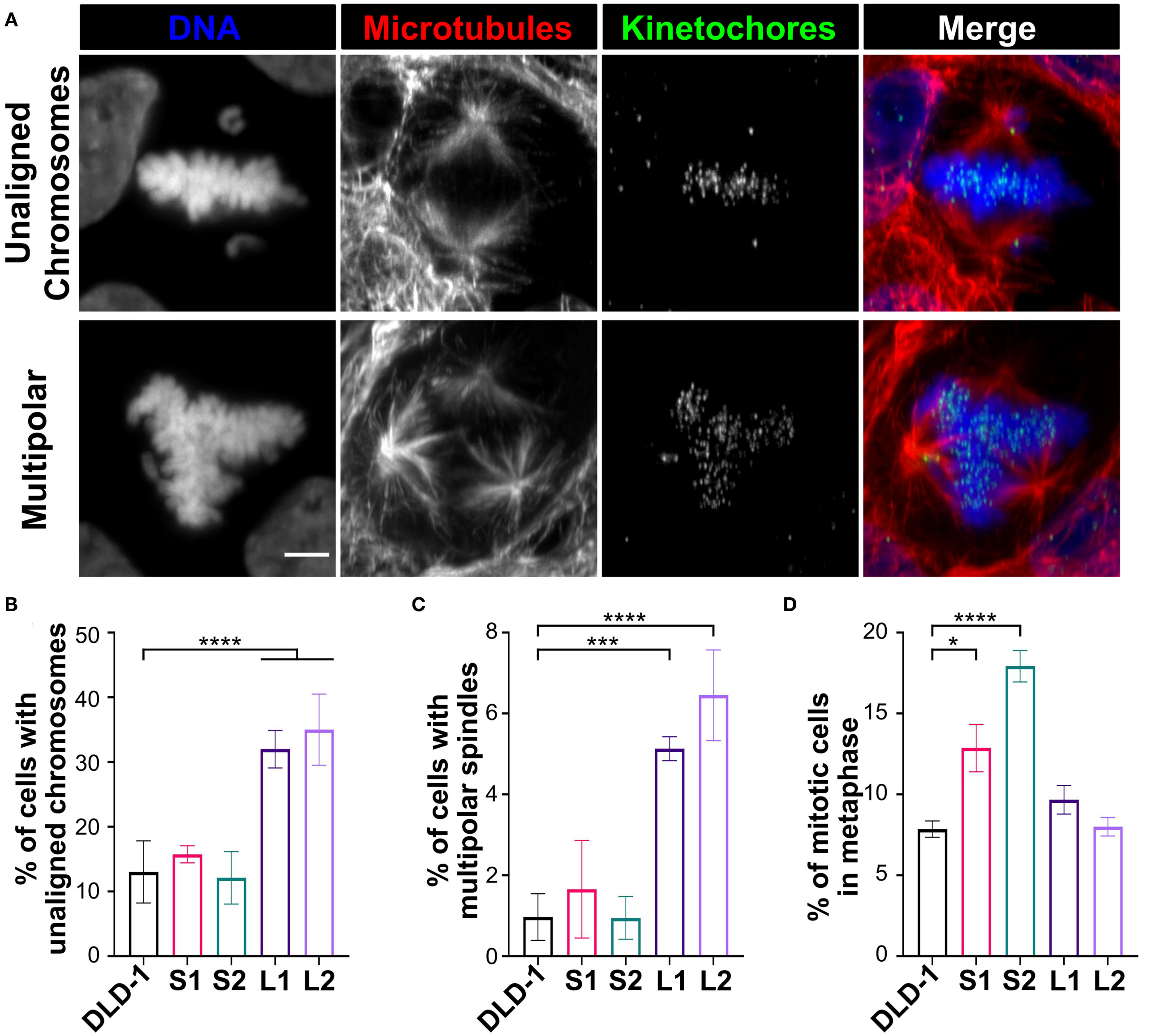


FIGURE 4 | Defects in prometaphase do not cause the mitotic delay in small 4N DLD-1 clones. **(A)** Examples of a cell with unaligned chromosomes (top) and a multipolar spindle (bottom). Scale bar, 5 μm. **(B–D)** Quantification of the percentage of late prometaphase-metaphase cells with **(B)** unaligned chromosome(s) and **(C)** multipolar spindles. **(D)** Quantification of the percentage of mitotic cells in metaphase. Graphs for panels **(B–D)** represent the mean ± SEM of three independent experiments where at least 100 cells were scored in each experiment ($n > 300$ for each cell line). $^{*}p < 0.05$, $^{***}p < 0.001$, $^{****}p < 0.0001$, when compared to parental DLD-1 cells using a two-sided Fisher's exact test.

in S1 and S2. When we examined metaphase duration, as expected, there was no significant difference between the 2N and 4N RPE-1 cells (**Supplementary Figure 3C**), whereas in each of the 4N DLD-1 clones, we observed an increase in metaphase duration compared to the parental cells. Consistent with our fixed-cell data, metaphase duration was the longest in S1 and S2 (**Figure 3D**). Specifically, whereas average metaphase duration was 17.1 min in the parental DLD-1 cells, it increased to 24.8 min in S2 and 35.4 min in S1 (**Figure 3D**). Altogether, these findings indicate that the mitotic delay in S1 and S2 results from a prolonged metaphase.

Differences in Spindle Dimensions and Cell Size Are Not Sufficient to Explain the Observed Differences in Metaphase Duration in DLD-1 Clones

Metaphase duration depends on the time required to silence the SAC after the last chromosome aligns at the metaphase plate (Rieder et al., 1994, 1995; Howell et al., 2000), and spindle and cell sizes have been proposed to impact the dynamics of SAC silencing (Chen and Liu, 2014, 2015, 2016). Thus, it is possible that the longer metaphase duration in our small 4N DLD-1 clones may stem from a size-dependent change in the time required

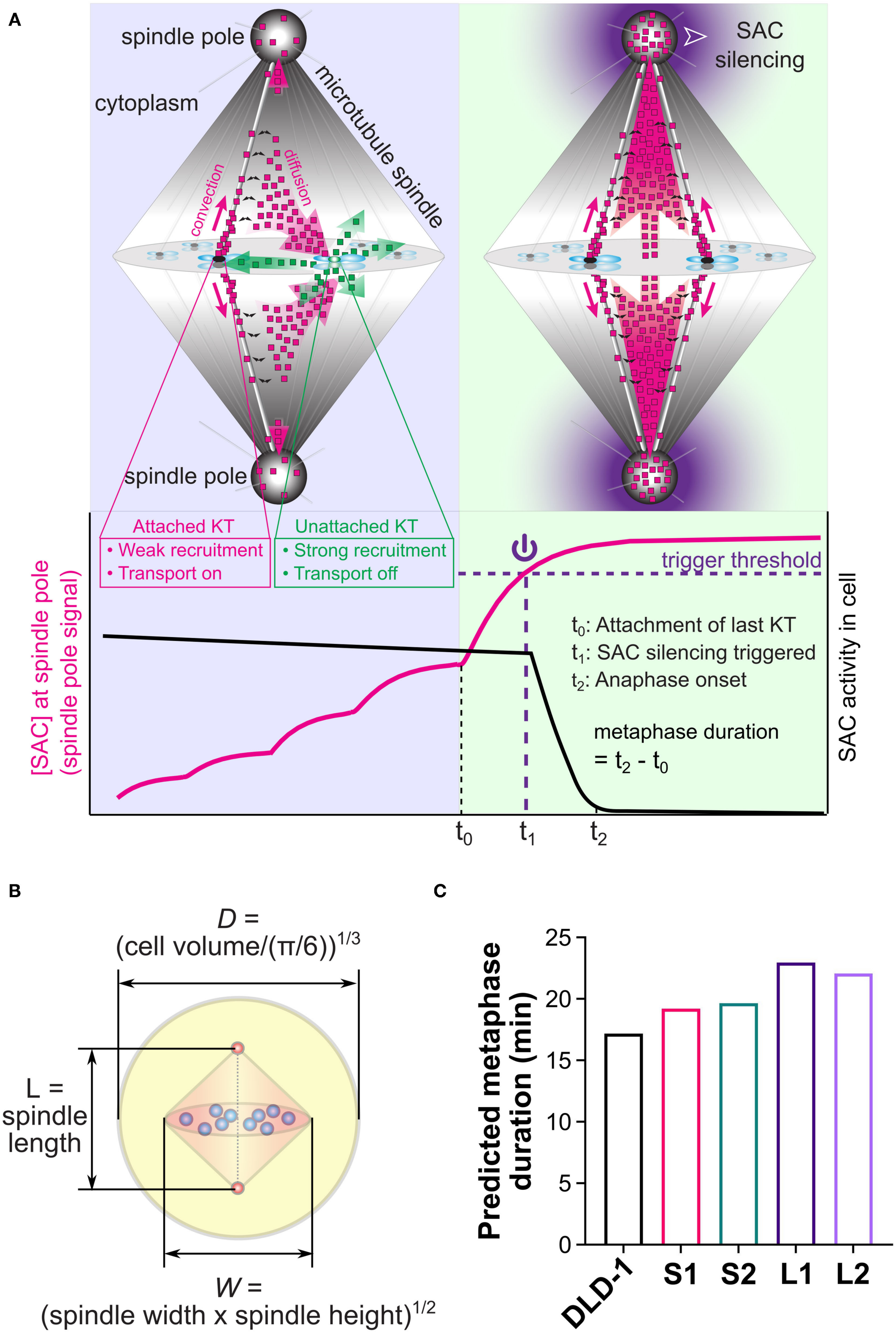


FIGURE 5 | Mathematical model for spatiotemporal regulation of SAC silencing. **(A)** Cartoon summary of the model. Upper panels illustrate the key model assumptions: (a) High phosphorylation level at unattached kinetochores (e.g., right kinetochore in the upper left panel) promotes recruitment of SAC proteins and

(Continued)

FIGURE 5 | inhibits their transport activity (shedding transport-inactive SAC proteins illustrated by green dots); (b) Low phosphorylation level at attached kinetochores (e.g., left kinetochore in the upper left panel and all kinetochores in the upper right panel) causes much weaker recruitment of SAC proteins and turns on their transport activity (shedding transport-active SAC proteins illustrated by magenta dots); (c) Transport-active SAC proteins (magenta dots) partially accumulate at the spindle pole, and partially escape from the poleward transport by dissociation from the microtubules (black double arrows), getting the chance to return to the kinetochores. Lower panel illustrates the model predicted dynamics of the SAC protein concentration at the spindle pole (magenta line) and overall SAC activity in the cell (black line). Through the spatiotemporal dynamics illustrated in the cartoons, SAC proteins accumulate at the spindle pole in a nonlinear fashion that depends on successive kinetochore attachments. The last kinetochore attachment causes a substantial jump in this spindle pole signal, which is assumed to trigger SAC silencing when crossing a threshold level (purple dashed line). The metaphase duration is the sum of the "triggering time" required for the spindle pole signal to reach the trigger threshold ($t_1 - t_0$) and the "propagation time" required for SAC silencing to propagate from the spindle pole throughout the cell ($t_2 - t_1$). **(B)** Cell and spindle dimensions in the model are calculated from experimental measurements. **(C)** Model predicted metaphase durations for the parental DLD-1 cells and the 4N DLD-1 clones using the cell and spindle dimensions shown in panel **(B)**. Values of the dimensions are given in **Table 1**. Panels **(A,B)** adapted with permission from (Chen and Liu, 2016).

for SAC silencing. To assess this possibility in explaining the intriguing relationship between cell/spindle sizes and metaphase duration observed in the 4N DLD-1 clones, we took advantage of a previous mathematical model (Chen and Liu, 2014) that addresses the spindle-mediated spatiotemporal regulation of SAC signals and associates the dynamics of SAC silencing with cell and spindle sizes (Chen and Liu, 2014, 2015). The model was built upon the phenomenon that SAC proteins are concentrated at the unattached kinetochores, but undergo continuous dynein-mediated transport from the attached kinetochores to the spindle pole (Howell et al., 2000, 2001; Wojcik et al., 2001; Basto et al., 2004; Griffis et al., 2007; Silva et al., 2014) (**Figure 5A**) (see Materials and Methods for basic model assumptions and **Supplementary Materials** for details and equations). The model predicts that such spatiotemporal regulation can lead to a nonlinear increase in the concentration of SAC proteins at the spindle pole, as more kinetochores within the cell achieve stable attachment to spindle microtubules (**Figure 5A**). Particularly, attachment of the last kinetochore can induce a drastic increase in the SAC concentration at the spindle pole. Such a dramatic increase can generate a noise-robust signal to silence the SAC after and only after all kinetochores are attached, if silencing is triggered by a proper threshold concentration of SAC proteins at the spindle pole (Chen and Liu, 2014) (**Figure 5A**). Thus, the model suggests that dynein-mediated poleward transport of SAC proteins could function to ensure robust, timely SAC silencing (Chen and Liu, 2014). In the model, the trigger signal for SAC silencing at the spindle pole could be any specific SAC protein, or any additional mitotic signaling protein undergoing the same spatiotemporal regulation as the SAC proteins, including cyclin B and APC/C (Acquaviva et al., 2004; Torres et al., 2010; Famulski et al., 2011). In the rest of the paper, we will refer to this putative SAC silencing trigger as the "spindle pole signal." Because the model treats the entire spindle, including kinetochores, microtubules and spindle poles, as an integrated mediator of the SAC silencing signal, it can predict how SAC silencing depends on spindle and cell sizes (Chen and Liu, 2016).

To assess the influence of size differences, we entered the median spindle and cell dimensions measured in the parental DLD-1 and 4N clones into the model (**Table 1**, **Figure 5B**), and calculated metaphase duration as the time delay between the last kinetochore attachment and SAC inactivation (**Figure 5A**, $t_2 - t_0$). We assumed the same values for the remaining model parameters across all cell clones, such as binding/unbinding constants and biochemical rate constants (**Supplementary Tables 1**, **2**). We found that the spindle and cell dimensions alone cannot explain the observed differences in metaphase duration in the 4N DLD-1 clones (**Figure 5C**). Specifically, the model predicts the two small 4N DLD-1 clones, S1 and S2, to have shorter metaphase durations than the two large 4N DLD-1 clones, L1 and L2. This is inconsistent with our experimental observations that metaphase is longer in S1 and S2 compared to L1 and L2. Furthermore, it is inconsistent with our finding that cyclin B persists in a larger fraction of metaphase cells in the small 4N DLD-1 clones compared to L1, L2, and parental DLD-1 cells (**Supplementary Figure 4**). Because cyclin B is degraded prior to anaphase onset (Chang et al., 2003), its persistence is indicative of a delay in SAC silencing.

Differences in Metaphase Duration Can Be Explained by Combined Differences in Cell Size, Spindle Size, Spindle Pole Size, and Spindle Microtubule Abundance

We next considered additional factors that can influence the dynamics of the spindle pole signal and consequently the predicted time for SAC silencing. The predicted metaphase duration reflects the sum of the "triggering time" and the "propagation time." The triggering time (**Figure 5A**, $t_1 - t_0$) is the time it takes the spindle pole signal to reach the trigger threshold after attachment of the last kinetochore. The triggering time also corresponds to the time period during which SAC silencing can be reversed by disrupting microtubule-kinetochore attachments (Dick and Gerlich, 2013). The propagation time (**Figure 5A**, $t_2 - t_1$) is the time it takes SAC silencing to propagate from the spindle pole throughout the cell. In the model, the propagation time largely depends on the biochemical pathway for SAC silencing, which does not differ significantly among clones derived from the same ancestor (**Supplementary Figure 5**), as they assume the same biochemical parameters. Instead, the difference in the predicted metaphase duration largely stems from the difference in the triggering time (**Supplementary Figure 5**), which heavily depends on the dynamics of the spindle pole signal, especially on how much the final spindle pole signal exceeds the threshold (**Figure 6A**). If the final spindle pole signal is very close to the threshold, it takes a considerably longer time for the signal to reach the threshold that is located near the plateau of the temporal curve (**Figure 6A**, purple solid line). Therefore, any

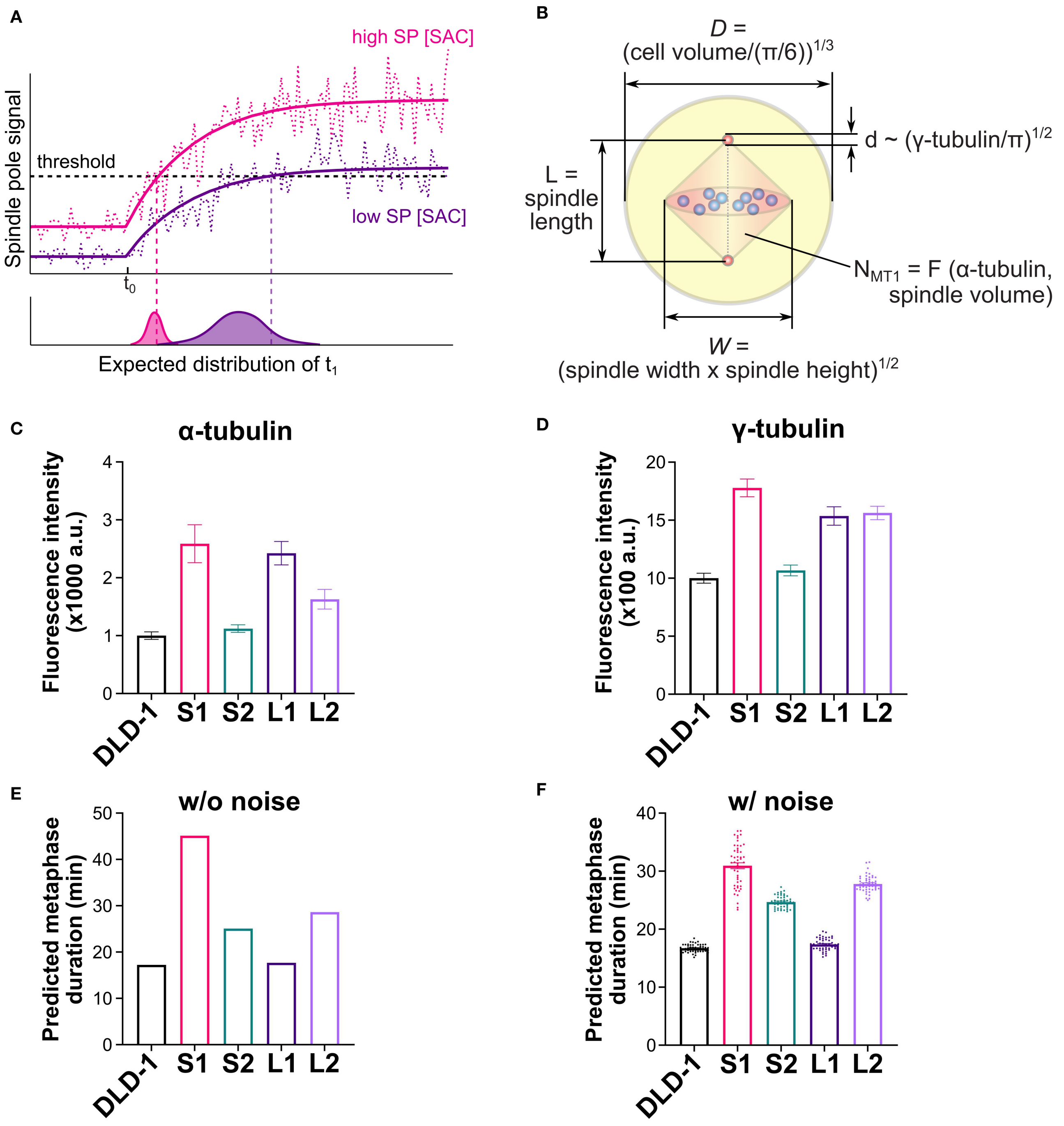


FIGURE 6 | Mathematical model explains observed metaphase durations when combining cell/spindle dimensions with tubulin measurements. **(A)** Effects of spindle pole signal level on predicted metaphase duration. A higher overall level of the spindle pole signal (magenta line) is expected to trigger SAC silencing earlier than a lower level does (purple line). Noise (dashed lines) is expected to expedite triggering of SAC silencing and cause variation in the triggering time, and both effects are expected to be more significant with lower spindle pole signal level. **(B)** Dimensions of cell, spindle and spindle pole, and abundance of microtubules calculated from experimental measurements (refer to the **Supplementary Materials** for the derivation of microtubule number, N_{MT1}). **(C)** Experimentally measured α-tubulin intensity in the parental DLD-1 cells and the 4N DLD-1 clones. **(D)** Experimentally measured γ-tubulin intensity in the parental DLD-1 cells and the 4N DLD-1 clones. **(E,F)** Metaphase durations predicted for the parental DLD-1 cells and the 4N DLD-1 clones using the dimensions shown in panel **(B)** without **(E)** and with **(F)** noise. Values of the dimensions are given in **Table 1**. Scatter dots in panel **(F)** show 50 individual simulation results and bars show the average.

factor that significantly impacts the spindle pole signal is expected to affect metaphase duration.

Aside from the spindle and cell dimensions, spindle microtubule abundance and spindle pole size can also strongly affect the spindle pole signal (**Supplementary Figure 6**) and therefore influence the triggering time. To obtain information on spindle microtubule abundance and spindle pole size, we quantified the fluorescence intensity of α- and γ-tubulin, respectively (**Figures 6C,D** and **Supplementary Figures S7A,B**). To obtain model parameters from these data, we converted the median α-tubulin signal intensity measured in each clone to the abundance of spindle microtubules in the model (**Figure 6B**, N_{MT1}) [**Table 1**; **Supplementary Materials**, Eq. (S2)]. We also took the median γ-tubulin signal intensity as a metric for spindle pole size and used it to derive the spindle pole diameters relative to the parental cells (**Figure 6B**, d) [**Table 1**; **Supplementary Materials**, Eq. (S1)]. When spindle microtubule abundance and spindle pole size were included, the model largely captured the experimentally observed relative metaphase durations in different DLD-1 clones, especially the observation that metaphase duration is much longer in S1 than all other clones (**Figure 6E**). The metaphase durations predicted for parental and 4N RPE-1 clones also matched the experimental data (**Supplementary Figure 7C**).

We further examined the variations in the predicted metaphase duration due to noise. When random noise in the spindle pole signal is considered (**Figure 6A**, dotted lines), SAC silencing has a chance to be triggered before the average signal reaches the trigger threshold. For cases with a high overall level of spindle pole signal, the threshold is located in the fast-rising region of the spindle pole signal and hence the triggering time is dominated by the rising dynamics and relatively smaller variation is expected (**Figure 6A**, lower plot, magenta distribution). For cases with a low overall level of spindle pole signal, as the signal slowly approaches the threshold in the near-plateau region, noise plays a much more important role, resulting in shorter and more stochastic triggering time (**Figure 6A**, lower plot, purple distribution). Indeed, when 20% noise was applied to the spindle pole signal in the model (see **Supplementary Materials**), the predicted metaphase duration in S1 was significantly lowered from 45 min in the deterministic simulation to 31 min average in the stochastic simulations with a wide distribution (**Figure 6F**). In contrast, the predicted metaphase durations for the parental DLD-1 cells and other 4N DLD-1 clones displayed average values close to the corresponding deterministic results and narrower distributions (**Figure 6F**). This is consistent with the experimental observation that metaphase duration in S1 was more stochastic (broader distribution) than in the parental cells and other 4N DLD-1 clones (**Figure 3D**). The parental and 4N RPE-1 clones are also predicted to have similar average metaphase duration as their corresponding deterministic predictions (**Supplementary Figures 7C,D**). In sum, our model shows that the difference in metaphase duration among the 4N clones can stem from the combined variations in cell size, spindle size, spindle pole size, and spindle microtubule abundance.

Changes in Spindle Shape Can Affect the Timing of SAC Silencing

In addition to displaying the longest metaphase duration, the S1 clone also displayed a unique spindle morphology, which we described in an earlier section as "compressed," given the substantial increase in spindle height, but not in spindle length, compared to the parental cells. We aimed to address whether this spindle morphology may contribute to the metaphase delay in these cells. Because S1 cells form nearly spherical mitotic cells, we reasoned that we may be able to alter spindle shape by limiting mitotic cell rounding. To this end, we treated S1 cells with the myosin inhibitor blebbistatin, since actomyosin contraction is important for mitotic cell rounding and spindle morphogenesis (Kunda et al., 2008; Kunda and Baum, 2009). Indeed, spindle size measurements (**Figures 7A–D**) showed that, at a low dose, blebbistatin treatment did not affect spindle width (**Figure 7A**) of S1 cells, but reduced spindle height by 10% and increased spindle length by 42% (**Figures 7B,C**) compared to untreated S1 cells. Using these size measurements in the model, we found that these changes in spindle dimensions decreased the predicted metaphase duration by 7 min (or 15%) (**Figure 7E**). To experimentally test this prediction, we quantified mitotic timing by time-lapse microscopy (**Figures 7F–H**) and found that mitotic duration was 11 min (or 18.6%) shorter in blebbistatin-treated S1 cells compared to untreated S1 cells (**Figure 7F**). Importantly, the decrease in mitotic duration could be explained by a significantly shorter metaphase duration (**Figure 7H**), in line with our model prediction. Altogether, these findings suggest that spindle shape plays an important role in timely SAC silencing and mitotic progression.

DISCUSSION

In this study, we generated tetraploid cell lines from RPE-1 and DLD-1 cells to examine the relationship between cell and spindle size scaling and mitotic progression. Our 4N DLD-1 clones could be separated into two groups (small and large) based on cell and nuclear volumes, supporting the notion that genome size does not necessarily set cell/nuclear size (Neumann and Nurse, 2007). This is also consistent with the observation that cells become progressively smaller during development, despite maintaining a constant genome size (Hara and Kimura, 2009). In contrast, we were not able to isolate a small 4N RPE-1 clone. The ability to suppress cell/nuclear size scaling in response to changes in DNA content may be specific to cancer cells and could have relevance for tumorigenesis. Indeed, DNA content and nuclear size do not always correlate in cancers (Wang et al., 1992). The mechanisms responsible for regulating and altering cell/nuclear size in normal and cancer cells warrant further study. Regardless of how much cell and nuclear sizes scaled, the N/C ratio was maintained in each cell type, which is consistent with many previous studies in which cell and nuclear sizes were examined in a variety of contexts (Conklin, 1912; Jorgensen et al., 2007; Neumann and Nurse, 2007).

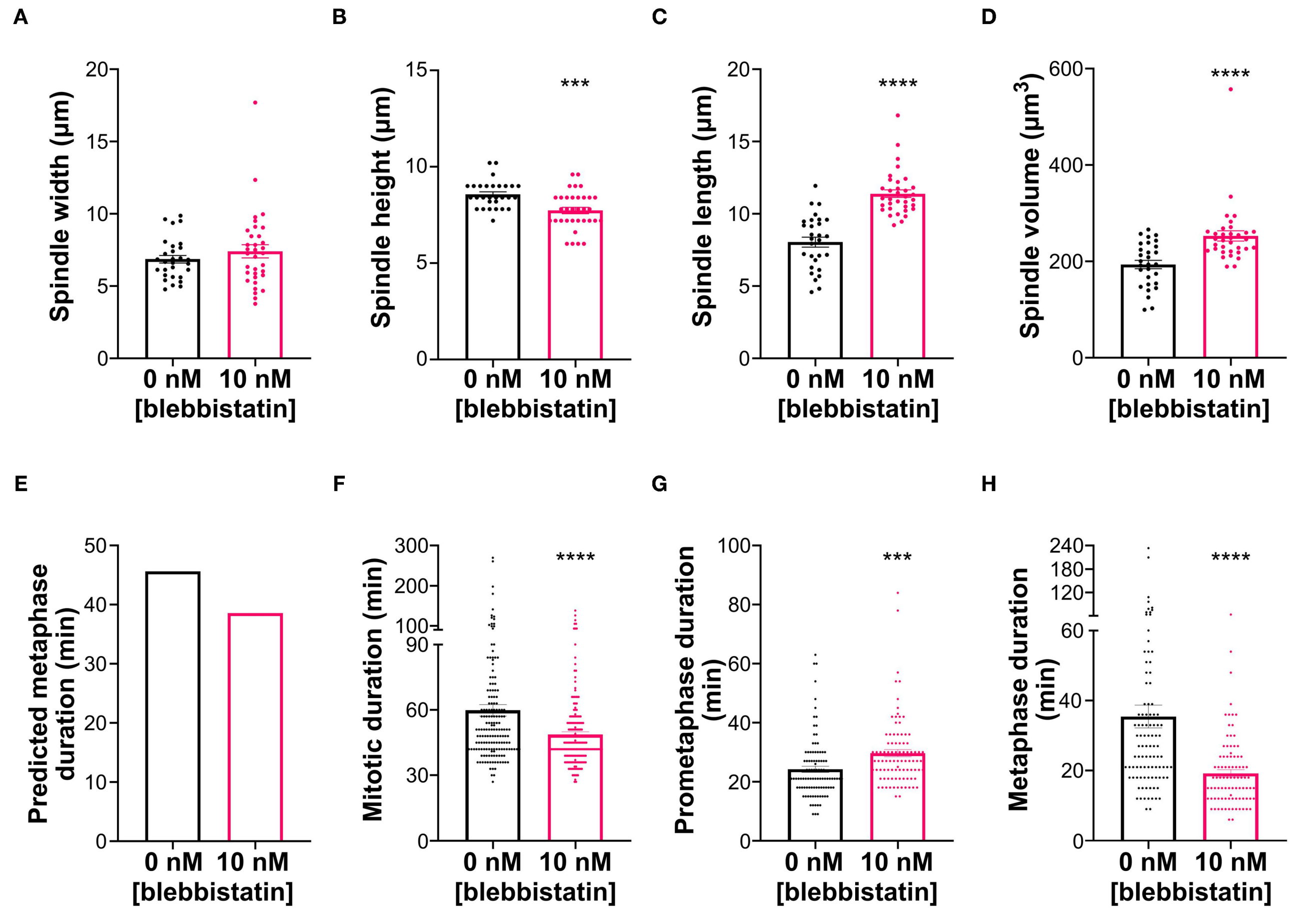


FIGURE 7 | Blebbistatin treatment modifies spindle dimensions and reduces metaphase duration in S1 cells. **(A–D)** Measurements of mitotic spindle width **(A)**, height **(B)**, length **(C)**, and volume **(D)** in untreated and blebbistatin-treated S1 cells ($n = 29$ and 36, respectively), reported as mean ± SEM with individual data points from three independent experiments. ***$p < 0.001$, ****$p < 0.0001$, when compared to untreated cells by Student's *t*-test. **(E)** Model-predicted metaphase duration using the median spindle dimensions in panels **(A–C)**. Cell size, spindle pole size, and spindle microtubule abundance were assumed to be the same between the control and treated groups. **(F–H)** Quantification of **(F)** mitotic duration in untreated and blebbistatin-treated S1 cells ($n = 167$ and 193, respectively), **(G)** prometaphase duration in untreated and blebbistatin-treated S1 cells ($n = 101$ and 92, respectively), and **(H)** metaphase duration in untreated and blebbistatin-treated S1 cells ($n = 101$ and 94, respectively). ***$p < 0.001$, ****$p < 0.0001$, when compared to untreated cells by Student's *t*-test.

Spindle Size Does Not Always Scale With Cell Size

Across many metazoan species and within developing organisms, spindle size typically scales with cell size, such that larger cells have longer spindles and vice versa (Brown et al., 2007; Loughlin et al., 2010, 2011; Good et al., 2013; Hazel et al., 2013; Reber et al., 2013; Wilbur and Heald, 2013; Crowder et al., 2015; Lacroix et al., 2018). Moreover, *Xenopus* egg extract spindles assembled in microfluidic devices were shown to adjust their size based on the available space (Good et al., 2013; Hazel et al., 2013). Our results, however, show that spindle geometry or size do not show an obvious correlation with cell size and that the relation between cell and spindle size is more complex than traditionally believed. Similar to a study in yeast (Storchová et al., 2006), spindle length did not increase with cell volume in the RPE-1 4N clone. Spindle width was similar in all of the 4N clones, despite differences in cell and spindle volumes, suggesting that the increase in chromosome number following tetraploidization may require a specific adjustment of spindle width. Moreover, width may only increase within certain constraints, given that excessively wide spindles can result in pole splitting and multipolarity (Dinarina et al., 2009; Lancaster et al., 2013). The lack of correlation between cell and spindle size was particularly evident in the small 4N DLD-1 clones, S1 and S2. Despite having similar cell and nuclear volumes, S1 and S2 displayed the smallest and largest spindle volumes, respectively, out of all the 4N DLD-1 clones, indicating that factors other than cell volume can determine spindle volume.

Of all the spindle dimensions, spindle height appeared to be critical for adjusting spindle volume in response to cell size variations. Indeed, the small 4N DLD-1 clones were characterized by spindles that were taller than the spindles in the large 4N clones. We found that inhibiting myosin with blebbistatin reduced spindle height in S1, indicating that actomyosin contraction during mitotic cell rounding assists in vertical expansion of the spindle. During mitotic cell rounding, which facilitates proper spindle formation (Kunda et al., 2008; Kunda and Baum, 2009), actomyosin contraction helps to counterbalance the outward force generated by osmotic pressure (Stewart et al., 2011). Since osmosis depends on the movement of ions and water across the cell membrane (Lang, 2007), the increased surface area-to-volume ratio of the small 4N DLD-1 clones may enable these cells to more readily regulate the osmotic gradients needed for mitotic rounding compared to the large 4N DLD-1 clones. Indeed, we found that mitotic cell rounding was most pronounced in S1 and S2, indicating that the efficiency of mitotic cell rounding may explain the relationship between cell size and spindle height in tetraploid cells. Mitotic rounding and cytoskeletal contraction were also shown to contribute to efficient mitotic spindle assembly and chromosome alignment (Lénárt et al., 2005; Lancaster et al., 2013; Booth et al., 2019), which can explain our observation that the small 4N clones had fewer defects in chromosome alignment compared to the large 4N clones. Thus, our findings suggest that the size of tetraploid cells can influence several aspects of mitosis, including mitotic rounding, spindle morphogenesis, and chromosome alignment.

Cell and Spindle Sizes Are Not the Only Determinants of SAC Silencing Timing

Previous studies have shown that SAC "strength" and mitotic duration increase as cell size decreases (Galli and Morgan, 2016; Gerhold et al., 2018). This is consistent with our observation that the small 4N clones displayed the longest mitotic durations among DLD-1 clones. Even with similar cell volumes, however, S1 and S2 had different mitotic and metaphase durations, indicating that cell size is not sufficient to explain the timing of SAC silencing. The spatiotemporal model for SAC silencing (Chen and Liu, 2014) we used here provides a theoretical framework to examine the combined effects of multiple parameters on SAC silencing time. For instance, among the cell and spindle dimensions, only spindle length differed significantly between S1 and S2, but the model predicts that the spindle pole signal is rather insensitive to spindle length (Chen and Liu, 2016). Hence, the difference in metaphase duration between S1 and S2 cannot be explained by differences in cell or spindle dimensions. However, the two small clones displayed particularly pronounced differences in their α- and γ-tubulin intensities, which translated to significantly higher spindle microtubule density and larger spindle poles in S1. According to the model, higher spindle microtubule density leads to stronger poleward flux of the SAC proteins, which increases the spindle pole signal and shortens metaphase duration. Meanwhile, larger spindle poles dilute the spindle pole signal and lengthen metaphase duration. Between S1 and S2, the impact of spindle pole size outcompeted the impact of spindle microtubule density, leading to a longer metaphase duration in S1. Unlike the comparison between S1 and S2, S1 and L1 displayed similar α- and γ-tubulin intensities. However, they showed different cell sizes and different spindle heights. Spindle height is particularly interesting because it has been shown that limiting cell height, and hence spindle height, through chemical or physical perturbations impairs mitotic progression (Dumont and Mitchison, 2009; Lancaster et al., 2013; Cattin et al., 2015). Our model suggested that the difference in metaphase duration between S1 and L1 stems from the larger cell size and smaller spindle height in L1, both of which increase the spindle pole signal and shorten metaphase duration. Consistent with this, when spindle height was experimentally reduced in S1 cells, metaphase duration was also reduced. Therefore, our data show that not only limiting spindle height, but also excessive spindle height (as that observed in S1 cells) can affect mitotic progression. Overall, our mathematical model showed that, although spindle scaling is important for robust SAC signaling (Chen and Liu, 2014, 2016), spindle size alone is not sufficient to predict the kinetics of SAC silencing. Instead, our data show that timing of SAC silencing can only be explained by the combined contributions of cell, spindle and pole sizes along with microtubule abundance.

Although the model explains most of the differences in metaphase duration among the parental cells and 4N DLD-1 clones, certain discrepancies remain between the model predictions and the experimental data. For example, the predicted metaphase duration in L2 is higher than all the other DLD-1 clones except for S1, which does not match the experimental data. Such a discrepancy could stem from additional factors that might differ among the clones, such as biochemical or transport activity of SAC proteins, which will require future investigations. The discrepancy could also be attributed to incompleteness of the model. The model has made a number of assumptions that yet await experimental verification. For example, SAC silencing is assumed to be triggered by a threshold concentration of a "spindle pole signal" that arrives by the dynein-mediated poleward transport. Although the dynein-mediated transport has been documented (Howell et al., 2000, 2001; Wojcik et al., 2001; Basto et al., 2004; Griffis et al., 2007; Silva et al., 2014), the identity of the signal, existence of the threshold, and the downstream pathway triggered by the threshold all need to be determined. Moreover, the biochemical part of the model is coarse-grained, resolving only a few key molecular players in the SAC mechanism. Nevertheless, using identical parameters for the biochemical circuits and protein transport across all the clones derived from the same ancestor, the model predictions provide insight on the relative differences in SAC silencing time among these clones, which is not attainable by statistical comparison of sizes alone.

Revisiting the Role of Tetraploidy and Spindle Architecture in Tumorigenesis

Tetraploidization has been proposed as an intermediate step during tumor progression (Storchova and Pellman, 2004; Ganem et al., 2007). This idea is supported by a number of observations, including the fact that nearly 40% of all tumors have likely undergone whole genome duplication during their clonal evolution (Zack et al., 2013) and that tetraploid, but not diploid, mammary epithelial cells could induce subcutaneous tumors in nude mice (Fujiwara et al., 2005). Newly formed tetraploid cells inherit extra centrosomes that promote spindle multipolarity (Chen et al., 2016; Baudoin et al., 2020), which was proposed as the mechanism underlying the tumorigenic potential of tetraploid cells (Storchova and Pellman, 2004). In our 4N clones, however, a majority of cells formed bipolar spindles with a normal centrosome number, consistent with other reports (Ganem et al., 2009; Godinho et al., 2014; Kuznetsova et al., 2015; Potapova et al., 2016; Viganó et al., 2018) and recent evidence that extra centrosomes are quickly lost through asymmetric centrosome clustering during the evolution of tetraploid cells (Baudoin et al., 2020). Our findings suggest that rather than multipolarity resulting from extra centrosomes, relatively subtle variations in spindle geometry, spindle pole size, microtubule abundance, and cell volume can affect mitotic progression and SAC silencing in tetraploid cells. Thus, even though spindle bipolarity may be preserved, other changes in cell size and spindle architecture—its geometry and composition—may promote mitotic dysfunction and contribute to the genomic instability caused by tetraploidy (Storchova and Kuffer, 2008; Dewhurst et al., 2014; Cohen-Sharir et al., 2020; Quinton et al., 2020).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Studies involving animal subjects: no animal studies are presented in this manuscript. Studies involving human subjects: no human studies are presented in this manuscript. Inclusion of identifiable human data: no potentially identifiable human images or data is presented in this study.

AUTHOR CONTRIBUTIONS

MB performed all the experiments, data acquisition, and data analysis. JC performed all the modeling work. DC provided resources and reagents. All authors contributed to conception of the study, data interpretation, figure preparation, and writing of the manuscript.

FUNDING

This work was partly funded by the Virginia Tech Center for Engineered Health, through seed funds to DC. Work in the Cimini lab was also supported by NSF grant MCB-1517506 and NIH grant R01GM140042 to DC. JC is supported by NIH (1R35GM138370) and startup funds from the Virginia Tech Department of Biological Sciences and College of Science.

ACKNOWLEDGMENTS

We would like to thank Dr. Josh Nicholson for generating two of the 4N clones used in this study. We further acknowledge members of the Cimini and Chen labs for helpful discussion. Finally, we would like to thank Dr. Silke Hauf for critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.596263/full#supplementary-material

REFERENCES

Acquaviva, C., Herzog, F., Kraft, C., and Pines, J. (2004). The anaphase promoting complex/cyclosome is recruited to centromeres by the spindle assembly checkpoint. *Nat. Cell Biol.* 6, 892–882. doi: 10.1038/ncb1167

Arnaoutov, A., Azuma, Y., Ribbeck, K., Joseph, J., Boyarchuk, Y., Karpova, T., et al. (2005). Crm1 is a mitotic effector of Ran-GTP in somatic cells. *Nat. Cell Biol.* 7, 626–632. doi: 10.1038/ncb1263

Basto, R., Scaerou, F., Mische, S., Wojcik, E., Lefebvre, C., Gomes, R., et al. (2004). In vivo dynamics of the rough deal checkpoint protein during Drosophila mitosis. *Curr. Biol.* 14, 56–61. doi: 10.1016/j.cub.2003.12.025

Baudoin, N. C., and Cimini, D. (2018). A guide to classifying mitotic stages and mitotic defects in fixed cells. *Chromosoma* 127, 215–227. doi: 10.1007/s00412-018-0660-2

Baudoin, N. C., Nicholson, J. M., Soto, K., Martin, O., Chen, J., and Cimini, D. (2020). Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. *Elife* 9:e54565. doi: 10.7554/eLife.54565.sa2

Bolte, S., and Cordelières, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* 224, 213–232. doi: 10.1111/j.1365-2818.2006.01706.x

Booth, A. J., Yue, Z., Eykelenboom, J. K., Stiff, T., Luxton, G. G., Hochegger, H., et al. (2019). Contractile acto-myosin network on nuclear envelope remnants positions human chromosomes for mitosis. *Elife* 8:e46902. doi: 10.7554/eLife.46902.040

Brewer, B. J., Chlebowicz-Sledziewska, E., and Fangman, W. L. (1984). Cell cycle phases in the unequal mother/daughter cell cycles of Saccharomyces cerevisiae. *Mol. Cell. Biol.* 4, 2529–2531. doi: 10.1128/MCB.4.11.2529

Brown, K. S., Blower, M. D., Maresca, T. J., Grammer, T. C., Harland, R. M., and Heald, R. (2007). Xenopus tropicalis egg extracts provide insight into scaling of the mitotic spindle. *J. Cell Biol.* 176, 765–770. doi: 10.1083/jcb.200610043

Cattin, C. J., Düggelin, M., Martinez-Martin, D., Gerber, C., Müller, D. J., and Stewart, M. P. (2015). Mechanical control of mitotic progression in single animal cells. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11258–11263. doi: 10.1073/pnas.1502029112

Chang, D. C., Xu, N., and Luo, K. Q. (2003). Degradation of cyclin B is required for the onset of anaphase in Mammalian cells. *J. Biol. Chem.* 278, 37865–37873. doi: 10.1074/jbc.M306376200

Chen, J., and Liu, J. (2014). Spatial-temporal model for silencing of the mitotic spindle assembly checkpoint. *Nat. Commun.* 5:4795. doi: 10.1038/ncomms5795

Chen, J., and Liu, J. (2015). Erroneous silencing of the mitotic checkpoint by aberrant spindle pole-kinetochore coordination. *Biophys. J.* 109, 2418–2435. doi: 10.1016/j.bpj.2015.10.024

Chen, J., and Liu, J. (2016). Spindle size scaling contributes to robust silencing of mitotic spindle assembly checkpoint. *Biophys. J.* 111, 1064–1077. doi: 10.1016/j.bpj.2016.07.039

Chen, S., Stout, J. R., Dharmaiah, S., Yde, S., Calvi, B. R., and Walczak, C. E. (2016). Transient endoreplication down-regulates the kinesin-14 HSET and contributes to genomic instability. *Mol. Biol. Cell* 27, 2911–2923. doi: 10.1091/mbc.E16-03-0159

Clute, P., and Pines, J. (1999). Temporal and spatial control of cyclin B1 destruction in metaphase. *Nat. Cell Biol.* 1, 82–87. doi: 10.1038/10049

Cohen-Sharir, Y., McFarland, J. M., Abdusamad, M., Marquis, C., Tang, H., Ippolito, M. R., et al. (2020). Selective vulnerability of aneuploid human cancer cells to inhibition of the spindle assembly checkpoint. *bioRxiv* [Preprint]. doi: 10.1101/2020.06.18.159038

Conklin, E. G. (1912). Cell size and nuclear size. *J. Exp. Zool.* 12, 1–98. doi: 10.1002/jez.1400120102

Crowder, M. E., Strzelecka, M., Wilbur, J. D., Good, M. C., von Dassow, G., and Heald, R. (2015). A comparative analysis of spindle morphometrics across metazoans. *Curr. Biol.* 25, 1542–1550. doi: 10.1016/j.cub.2015.04.036

Dewhurst, S. M., McGranahan, N., Burrell, R. A., Rowan, A. J., Grönroos, E., Endesfelder, D., et al. (2014). Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Dis.* 4, 175–185. doi: 10.1158/2159-8290.CD-13-0285

Dick, A. E., and Gerlich, D. W. (2013). Kinetic framework of spindle assembly checkpoint signalling. *Nat. Cell Biol.* 15, 1370–1377. doi: 10.1038/ncb2842

Dinarina, A., Pugieux, C., Corral, M. M., Loose, M., Spatz, J., Karsenti, E., et al. (2009). Chromatin shapes the mitotic spindle. *Cell* 138, 502–513. doi: 10.1016/j.cell.2009.05.027

Ditchfield, C., Johnson, V. L., Tighe, A., Ellston, R., Haworth, C., Johnson, T., et al. (2003). Aurora B couples chromosome alignment with anaphase by targeting BubR1, Mad2, and Cenp-E to kinetochores. *J. Cell Biol.* 161, 267–280. doi: 10.1083/jcb.200208091

Dumont, S., and Mitchison, T. J. (2009). Compression regulates mitotic spindle length by a mechanochemical switch at the poles. *Curr. Biol.* 19, 1086–1095. doi: 10.1016/j.cub.2009.05.056

Edgar, B. A., and Orr-Weaver, T. L. (2001). Endoreplication cell cycles: more for less. *Cell* 105, 297–306. doi: 10.1016/S0092-8674(01)00334-8

Etemad, B., and Kops, G. J. (2016). Attachment issues: kinetochore transformations and spindle checkpoint silencing. *Curr. Opin. Cell Biol.* 39, 101–108. doi: 10.1016/j.ceb.2016.02.016

Famulski, J. K., Vos, L. J., Rattner, J. B., and Chan, G. K. (2011). Dynein/Dynactin-mediated transport of kinetochore components off kinetochores and onto spindle poles induced by nordihydroguaiaretic acid. *PLoS One* 6:e16494. doi: 10.1371/journal.pone.0016494

Foley, E. A., and Kapoor, T. M. (2013). Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore. *Nat. Rev. Mol. Cell Biol.* 14, 25–37. doi: 10.1038/nrm3494

Fujiwara, T., Bandi, M., Nitta, M., Ivanova, E. V., Bronson, R. T., and Pellman, D. (2005). Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* 437:1043. doi: 10.1038/nature04217

Galli, M., and Morgan, D. O. (2016). Cell size determines the strength of the spindle assembly checkpoint during embryonic development. *Dev. Cell* 36, 344–352. doi: 10.1016/j.devcel.2016.01.003

Ganem, N. J., Godinho, S. A., and Pellman, D. (2009). A mechanism linking extra centrosomes to chromosomal instability. *Nature* 460:278. doi: 10.1038/nature08136

Ganem, N. J., Storchova, Z., and Pellman, D. (2007). Tetraploidy, aneuploidy and cancer. *Curr. Opin. Genet. Dev.* 17, 157–162. doi: 10.1016/j.gde.2007.02.011

Gassmann, R., Holland, A. J., Varma, D., Wan, X., Civril, F., Cleveland, D. W., et al. (2010). Removal of Spindly from microtubule-attached kinetochores controls spindle checkpoint silencing in human cells. *Genes Dev.* 24, 957–971. doi: 10.1101/gad.1886810

Gelens, L., Qian, J., Bollen, M., and Saurin, A. T. (2018). The importance of kinase-phosphatase integration: lessons from mitosis. *Trends cell Biol.* 28, 6–21. doi: 10.1016/j.tcb.2017.09.005

Gerhold, A. R., Poupart, V., Labbé, J.-C., and Maddox, P. S. (2018). Spindle assembly checkpoint strength is linked to cell fate in the *Caenorhabditis elegans* embryo. *Mol. Biol. Cell* 29, 1435–1448. doi: 10.1091/mbc.E18-04-0215

Gerhold, A. R., Ryan, J., Vallée-Trudeau, J.-N., Dorn, J. F., Labbé, J.-C., and Maddox, P. S. (2015). Investigating the regulation of stem and progenitor cell mitotic progression by *in situ* imaging. *Curr. Biol.* 25, 1123–1134. doi: 10.1016/j.cub.2015.02.054

Gillooly, J. F., Hein, A., and Damiani, R. (2015). Nuclear DNA content varies with cell size across human cell types. *Cold Spring Harbor Perspect. Biol.* 7:a019091. doi: 10.1101/cshperspect.a019091

Godinho, S. A., Picone, R., Burute, M., Dagher, R., Su, Y., Leung, C. T., et al. (2014). Oncogene-like induction of cellular invasion from centrosome amplification. *Nature* 510, 167–171. doi: 10.1038/nature13277

Good, M. C., Vahey, M. D., Skandarajah, A., Fletcher, D. A., and Heald, R. (2013). Cytoplasmic volume modulates spindle size during embryogenesis. *Science* 342, 856–860. doi: 10.1126/science.1243147

Gorbsky, G. J., Chen, R.-H., and Murray, A. W. (1998). Microinjection of antibody to Mad2 protein into mammalian cells in mitosis induces premature anaphase. *J. Cell Biol.* 141, 1193–1205. doi: 10.1083/jcb.141.5.1193

Gregory, T. R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev.* 76, 65–101. doi: 10.1017/S1464793100005595

Griffis, E. R., Stuurman, N., and Vale, R. D. (2007). Spindly, a novel protein essential for silencing the spindle assembly checkpoint, recruits dynein to the kinetochore. *J. Cell Biol.* 177, 1005–1015. doi: 10.1083/jcb.200702062

Hagting, A., den Elzen, N., Vodermaier, H. C., Waizenegger, I. C., Peters, J.-M., and Pines, J. (2002). Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. *J. Cell Biol.* 157, 1125–1137. doi: 10.1083/jcb.200111001

Hara, Y., and Kimura, A. (2009). Cell-size-dependent spindle elongation in the *Caenorhabditis elegans* early embryo. *Curr. Biol.* 19, 1549–1554. doi: 10.1016/j.cub.2009.07.050

Hauf, S., Cole, R. W., LaTerra, S., Zimmer, C., Schnapp, G., Walter, R., et al. (2003). The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* 161, 281–294. doi: 10.1083/jcb.200208092

Hazel, J., Krutkramelis, K., Mooney, P., Tomschik, M., Gerow, K., Oakey, J., et al. (2013). Changes in cytoplasmic volume are sufficient to drive spindle scaling. *Science* 342, 853–856. doi: 10.1126/science.1243110

Heald, R., and Gibeaux, R. (2018). Subcellular scaling: does size matter for cell division? *Curr. Opin. Cell Biol.* 52, 88–95. doi: 10.1016/j.ceb.2018.02.009

Heasley, L. R., Markus, S. M., and DeLuca, J. G. (2017). "Wait anaphase" signals are not confined to the mitotic spindle. *Mol. Biol. Cell* 28, 1186–1194. doi: 10.1091/mbc.e17-01-0036

Howell, B., McEwen, B., Canman, J., Hoffman, D., Farrar, E., Rieder, C., et al. (2001). Cytoplasmic dynein/dynactin drives kinetochore protein transport to the spindle poles and has a role in mitotic spindle checkpoint inactivation. *J. Cell Biol.* 155, 1159–1172. doi: 10.1083/jcb.200105093

Howell, B. J., Hoffman, D., Fang, G., Murray, A., and Salmon, E. (2000). Visualization of Mad2 dynamics at kinetochores, along spindle fibers, and at spindle poles in living cells. *J. Cell Biol.* 150, 1233–1250. doi: 10.1083/jcb.150.6.1233

Jorgensen, P., Edgington, N. P., Schneider, B. L., Rupeš, I., Tyers, M., and Futcher, B. (2007). The size of the nucleus increases as yeast cells grow. *Mol. Biol. Cell* 18, 3523–3532. doi: 10.1091/mbc.e06-10-0973

Khodjakov, A., and Rieder, C. L. (2009). The nature of cell-cycle checkpoints: facts and fallacies. *J. Biol.* 8:88. doi: 10.1186/jbiol195

King, S. J., and Schroer, T. A. (2000). Dynactin increases the processivity of the cytoplasmic dynein motor. *Nat. Cell Biol.* 2, 20–24. doi: 10.1038/71338

Krüger, L. K., Sanchez, J.-L., Paoletti, A., and Tran, P. T. (2019). Kinesin-6 regulates cell-size-dependent spindle elongation velocity to keep mitosis duration constant in fission yeast. *Elife* 8:e42182. doi: 10.7554/eLife.42182.037

Kunda, P., and Baum, B. (2009). The actin cytoskeleton in spindle assembly and positioning. *Trends Cell Biol.* 19, 174–179. doi: 10.1016/j.tcb.2009.01.006

Kunda, P., Pelling, A. E., Liu, T., and Baum, B. (2008). Moesin controls cortical rigidity, cell rounding, and spindle morphogenesis during mitosis. *Curr. Biol.* 18, 91–101. doi: 10.1016/j.cub.2007.12.051

Kuznetsova, A. Y., Seget, K., Moeller, G. K., de Pagter, M. S., de Roos, J. A., Dürrbaum, M., et al. (2015). Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle* 14, 2810–2820. doi: 10.1080/15384101.2015.1068482

Kyogoku, H., and Kitajima, T. S. (2017). Large cytoplasm is linked to the error-prone nature of oocytes. *Dev. Cell* 41, 287–298.e284. doi: 10.1016/j.devcel.2017.04.009

Lacroix, B., Letort, G., Pitayu, L., Sallé, J., Stefanutti, M., Maton, G., et al. (2018). Microtubule dynamics scale with cell size to set spindle length and assembly timing. *Dev. Cell* 45, 496–511.e496. doi: 10.1016/j.devcel.2018.04.022

Lancaster, O. M., Le Berre, M., Dimitracopoulos, A., Bonazzi, D., Zlotek-Zlotkiewicz, E., Picone, R., et al. (2013). Mitotic rounding alters cell geometry to ensure efficient bipolar spindle formation. *Dev. Cell* 25, 270–283. doi: 10.1016/j.devcel.2013.03.014

Lang, F. (2007). Mechanisms and significance of cell volume regulation. *J. Am. Coll. Nutr.* 26, 613S−623S. doi: 10.1080/07315724.2007.10719667

Leitao, R. M., and Kellogg, D. R. (2017). The duration of mitosis and daughter cell size are modulated by nutrients in budding yeast. *J. Cell Biol.* 216, 3463–3470. doi: 10.1083/jcb.201609114

Lénárt, P., Bacher, C. P., Daigle, N., Hand, A. R., Eils, R., Terasaki, M., et al. (2005). A contractile nuclear actin network drives chromosome congression in oocytes. *Nature* 436, 812–818. doi: 10.1038/nature03810

Liu, D., Vader, G., Vromans, M. J., Lampson, M. A., and Lens, S. M. (2009). Sensing chromosome bi-orientation by spatial separation of aurora B kinase from kinetochore substrates. *Science* 323, 1350–1353. doi: 10.1126/science.1167000

Loughlin, R., Heald, R., and Nédélec, F. (2010). A computational model predicts Xenopus meiotic spindle organization. *J. Cell Biol.* 191, 1239–1249. doi: 10.1083/jcb.201006076

Loughlin, R., Wilbur, J. D., McNally, F. J., Nédélec, F. J., and Heald, R. (2011). Katanin contributes to interspecies spindle length scaling in Xenopus. *Cell* 147, 1397–1407. doi: 10.1016/j.cell.2011.11.014

Maresca, T. J., and Salmon, E. D. (2009). Intrakinetochore stretch is associated with changes in kinetochore phosphorylation and spindle assembly checkpoint activity. *J. Cell Biol.* 184, 373–381. doi: 10.1083/jcb.200808130

Matson, D. R., and Stukenberg, P. T. (2014). CENP-I and Aurora B act as a molecular switch that ties RZZ/Mad1 recruitment to kinetochore attachment status. *J. Cell Biol.* 205, 541–554. doi: 10.1083/jcb.201307137

Mayer, V. W., Goin, C. J., Arras, C. A., and Taylor-Mayer, R. E. (1992). Comparison of chemically induced chromosome loss in a diploid, triploid, and tetraploid strain of *Saccharomyces cerevisiae*. *Mutat. Res. Genet. Toxicol.* 279, 41–48. doi: 10.1016/0165-1218(92)90264-Z

McCleland, M. L., Farrell, J. A., and O'Farrell, P. H. (2009). Influence of cyclin type and dose on mitotic entry and progression in the early Drosophila embryo. *J. Cell Biol.* 184, 639–646. doi: 10.1083/jcb.200810012

Meraldi, P., Draviam, V. M., and Sorger, P. K. (2004). Timing and checkpoints in the regulation of mitotic progression. *Dev. Cell* 7, 45–60. doi: 10.1016/j.devcel.2004.06.006

Mogilner, A., and Craig, E. (2010). Towards a quantitative understanding of mitotic spindle assembly and mechanics. *J. Cell Sci.* 123, 3435–3445. doi: 10.1242/jcs.062208

Mortimer, R. K. (1958). Radiobiological and genetic studies on a polyploid series (haploid to hexaploid) of *Saccharomyces cerevisiae*. *Radiat. Res.* 9, 312–326. doi: 10.2307/3570795

Moura, M., Osswald, M., Leça, N., Barbosa, J., Pereira, A. J., Maiato, H., et al. (2017). Protein phosphatase 1 inactivates Mps1 to ensure efficient spindle assembly checkpoint silencing. *Elife* 6:e25366. doi: 10.7554/eLife.25366.040

Musacchio, A. (2011). Spindle assembly checkpoint: the third decade. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 3595–3604. doi: 10.1098/rstb.2011.0072

Musacchio, A. (2015). The molecular biology of spindle assembly checkpoint signaling dynamics. *Curr. Biol.* 25, R1002–R1018. doi: 10.1016/j.cub.2015.08.051

Neumann, F. R., and Nurse, P. (2007). Nuclear size control in fission yeast. *J. Cell Biol.* 179, 593–600. doi: 10.1083/jcb.200708054

Nijenhuis, W., Vallardi, G., Teixeira, A., Kops, G. J., and Saurin, A. T. (2014). Negative feedback at kinetochores underlies a responsive spindle checkpoint signal. *Nat. Cell Biol.* 16, 1257–1264. doi: 10.1038/ncb3065

Paim, L. M. G., and FitzHarris, G. (2019). Tetraploidy causes chromosomal instability in acentriolar mouse embryos. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-12772-8

Pereira, A. J., and Maiato, H. (2012). Maturation of the kinetochore-microtubule interface and the meaning of metaphase. *Chromosome Res.* 20, 563–577. doi: 10.1007/s10577-012-9298-8

Potapova, T., and Gorbsky, G. J. (2017). The consequences of chromosome segregation errors in mitosis and meiosis. *Biology* 6:12. doi: 10.3390/biology6010012

Potapova, T. A., Seidel, C. W., Box, A. C., Rancati, G., and Li, R. (2016). Transcriptome analysis of tetraploid cells identifies cyclin D2 as a facilitator of adaptation to genome doubling in the presence of p53. *Mol. Biol. Cell* 27, 3065–3084. doi: 10.1091/mbc.e16-05-0268

Quinton, R. J., DiDomizio, A., Vittoria, M. A., Ticas, C. J., Patel, S., Koga, Y., et al. (2020). Whole genome doubling confers unique genetic vulnerabilities on tumor cells. *bioRxiv* [Preprint]. doi: 10.1101/2020.06.18.159095

Rasband, W. S. (2011). *National Institutes of Health*. Bethesda, MD. Available online at: http://imagej. nih. gov/ij/.

Reber, S. B., Baumgart, J., Widlund, P. O., Pozniakovsky, A., Howard, J., Hyman, A. A., et al. (2013). XMAP215 activity sets spindle length by controlling the total mass of spindle microtubules. *Nat. Cell Biol.* 15, 1116–1122. doi: 10.1038/ncb2834

Reck-Peterson, S. L., Yildiz, A., Carter, A. P., Gennerich, A., Zhang, N., and Vale, R. D. (2006). Single-molecule analysis of dynein processivity and stepping behavior. *Cell* 126, 335–348. doi: 10.1016/j.cell.2006.05.046

Rieder, C. L., Cole, R. W., Khodjakov, A., and Sluder, G. (1995). The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J. Cell Biol.* 130, 941–948. doi: 10.1083/jcb.130.4.941

Rieder, C. L., and Maiato, H. (2004). Stuck in division or passing through: what happens when cells cannot satisfy the spindle assembly checkpoint. *Dev. Cell* 7, 637–651. doi: 10.1016/j.devcel.2004.09.002

Rieder, C. L., Schultz, A., Cole, R., and Sluder, G. (1994). Anaphase onset in vertebrate somatic cells is controlled by a checkpoint that monitors sister kinetochore attachment to the spindle. *J. Cell Biol.* 127, 1301–1310. doi: 10.1083/jcb.127.5.1301

Sacristan, C., and Kops, G. J. (2015). Joined at the hip: kinetochores, microtubules, and spindle assembly checkpoint signaling. *Trends Cell Biol.* 25, 21–28. doi: 10.1016/j.tcb.2014.08.006

Saurin, A. T. (2018). Kinase and phosphatase cross-talk at the kinetochore. *Front. Cell Dev. Biol.* 6:62. doi: 10.3389/fcell.2018.00062

Shah, J. V., Botvinick, E., Bonday, Z., Furnari, F., Berns, M., and Cleveland, D. W. (2004). Dynamics of centromere and kinetochore proteins: implications for checkpoint signaling and silencing. *Curr. Biol.* 14, 942–952. doi: 10.1016/S0960-9822(04)00381-1

Sikora-Polaczek, M., Hupalowska, A., Polanski, Z., Kubiak, J. Z., and Ciemerych, M. A. (2006). The first mitosis of the mouse embryo is prolonged by transitional metaphase arrest. *Biol. Reprod.* 74, 734–743. doi: 10.1095/biolreprod.105.047092

Silkworth, W. T., Nardi, I. K., Scholl, L. M., and Cimini, D. (2009). Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PLoS One* 4:e6564. doi: 10.1371/journal.pone.0006564

Silva, P. M., Reis, R. M., Bolanos-Garcia, V. M., Florindo, C., Tavares, A. A., and Bousbaa, H. (2014). Dynein-dependent transport of spindle assembly checkpoint proteins off kinetochores toward spindle poles. *FEBS Lett.* 588, 3265–3273. doi: 10.1016/j.febslet.2014.07.011

Sivakumar, S., Daum, J. R., Tipton, A. R., Rankin, S., and Gorbsky, G. J. (2014). The spindle and kinetochore-associated (Ska) complex enhances binding of the anaphase-promoting complex/cyclosome (APC/C) to chromosomes and promotes mitotic exit. *Mol. Biol. Cell* 25, 594–605. doi: 10.1091/mbc.e13-07-0421

Smith, R. J., Cordeiro, M. H., Davey, N. E., Vallardi, G., Ciliberto, A., Gross, F., et al. (2019). PP1 and PP2A use opposite phospho-dependencies to control distinct processes at the kinetochore. *Cell Rep.* 28, 2206–2219.e2208. doi: 10.1016/j.celrep.2019.07.067

Stewart, M. P., Helenius, J., Toyoda, Y., Ramanathan, S. P., Muller, D. J., and Hyman, A. A. (2011). Hydrostatic pressure and the actomyosin cortex drive mitotic cell rounding. *Nature* 469, 226–230. doi: 10.1038/nature09642

Storchová, Z., Breneman, A., Cande, J., Dunn, J., Burbank, K., O'Toole, E., et al. (2006). Genome-wide genetic analysis of polyploidy in yeast. *Nature* 443:541. doi: 10.1038/nature05178

Storchova, Z., and Kuffer, C. (2008). The consequences of tetraploidy and aneuploidy. *J. Cell Sci.* 121, 3859–3866. doi: 10.1242/jcs.039537

Storchova, Z., and Pellman, D. (2004). From polyploidy to aneuploidy, genome instability and cancer. *Nat. Rev. Mol. Cell Biol.* 5, 45–54. doi: 10.1038/nrm1276

Tauchman, E. C., Boehm, F. J., and DeLuca, J. G. (2015). Stable kinetochore-microtubule attachment is sufficient to silence the spindle assembly checkpoint in human cells. *Nat. Commun.* 6, 1–9. doi: 10.1038/ncomms10036

Taylor, S. S., Scott, M. I., and Holland, A. J. (2004). The spindle checkpoint: a quality control mechanism which ensures accurate chromosome segregation. *Chromosome Res.* 12, 599–616. doi: 10.1023/B:CHRO.0000036610.78380.51

Torres, J. Z., Ban, K. H., and Jackson, P. K. (2010). A specific form of phospho protein phosphatase 2 regulates anaphase-promoting complex/cyclosome association with spindle poles. *Mol. Biol. Cell* 21, 897–904. doi: 10.1091/mbc.e09-07-0598

Tyson, J. J., Chen, K. C., and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15, 221–231. doi: 10.1016/S0955-0674(03)00017-6

Vassilev, L. T. (2006). Cell cycle synchronization at the G2/M phase border by reversible inhibition of CDK1. *Cell Cycle* 5, 2555–2556. doi: 10.4161/cc.5.22.3463

Vázquez-Diez, C., Paim, L. M. G., and FitzHarris, G. (2019). Cell-size-independent spindle checkpoint failure underlies chromosome segregation error in mouse embryos. *Curr. Biol.* 29, 865–873.e863. doi: 10.1016/j.cub.2018.12.042

Viganó, C., von Schubert, C., Ahrné, E., Schmidt, A., Lorber, T., Bubendorf, L., et al. (2018). Quantitative proteomic and phosphoproteomic comparison of human colon cancer DLD-1 cells differing in ploidy and chromosome stability. *Mol. Biol. Cell* 29, 1031–1047. doi: 10.1091/mbc.E17-10-0577

Wang, N., Stenkvist, B., and Tribukait, B. (1992). Morphometry of nuclei of the normal and malignant prostate in relation to DNA ploidy. *Anal. Quant. Cytol. Histol.* 14, 210–216.

Welburn, J. P., Vleugel, M., Liu, D., Yates, J. R. 3rd, Lampson, M. A., Fukagawa, T., et al. (2010). Aurora B phosphorylates spatially distinct targets to differentially regulate the kinetochore-microtubule interface. *Mol. Cell* 38, 383–392. doi: 10.1016/j.molcel.2010.02.034

Whyte, J., Bader, J. R., Tauhata, S. B., Raycroft, M., Hornick, J., Pfister, K. K., et al. (2008). Phosphorylation regulates targeting of cytoplasmic dynein to kinetochores during mitosis. *J. Cell Biol.* 183, 819–834. doi: 10.1083/jcb.200804114

Wilbur, J. D., and Heald, R. (2013). Mitotic spindle scaling during Xenopus development by kif2a and importin α. *Elife* 2:e00290. doi: 10.7554/eLife.00290

Wojcik, E., Basto, R., Serr, M., Scaerou, F., Karess, R., and Hays, T. (2001). Kinetochore dynein: its dynamics and role in the transport of the Rough deal checkpoint protein. *Nat. Cell Biol.* 3, 1001–1007. doi: 10.1038/ncb1101-1001

Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* 45, 1134–1140. doi: 10.1038/ng.2760

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership