Computer aided interactive gene network simulations including stochastic molecular kinetics and noise

Johannes Hettich and J. Christof M. Gebhardt[#]
Institute of Biophysics, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm
[#]To whom correspondence should be addressed: christof.gebhardt@uni-ulm.de

Abstract

Recent next generation sequencing and single molecule methodologies provide functional insights into gene regulatory networks beyond Boolean interactions, including the number, binding modes and kinetic rates of transcription factors and the kinetics of gene bursting. We report CaiNet, a fast computer aided interactive network simulation environment to set up and simulate arbitrary gene regulatory networks. CaiNet automatically compiles a network laid out in a graphical user interface into a fast hybrid stochastic-deterministic simulation framework without further mathematical knowledge or input by the user. Stochastic noise can optionally be omitted for simplified deterministic solutions. We validate CaiNet by comparison with Gillespie simulations using an auto feedback motive. We apply CaiNet to the circadian clock and find that temporally modulated external input signals allow regulating the periodicity of oscillations in a nested network topology. We further use CaiNet to simulate the temporal behavior of the pluripotency network, using published kinetic parameters where possible, as it transits from the naïve state to germ layer lineages upon changes in signal inputs.

Introduction

Dynamics and progression of many fundamental processes in cells and organisms, including metabolism [1, 2] the cell cycle [3], the circadian clock [4], differentiation [5-7] and development [8] are governed by gene regulatory networks. These networks generally control the activity of genes by regulatory motifs such as feedback or feed forward loops [9, 10], which ensure spatially and temporally controlled gene expression. The structure of gene regulatory networks commonly is inferred from large scale gene expression [11, 12] and transcription factor target site [13] analysis by comparing different network conditions, e.g. at different time points or before and after knockdown of genes. Regulatory networks have been investigated with logic[1, 14-16], continuous[17] and inference approaches dedicated to gene regulatory networks[18]. Such models yield basic dynamic descriptions of networks, yet neglect details associated with the stochastic behavior of biomolecules on the single molecule level such as noise in gene expression[19-21].

Since recent years, novel experimental approaches are emerging that enable obtaining detailed molecular and kinetic information about biological processes underlying gene regulatory networks. They reveal whether transcription factors bind simultaneously to promoters or enhancers[22], the kinetics of gene bursting [23-25], elongation [26-28], transcription [29-31] and translation rates [32, 33], mRNA [34] and protein abundances [35-38], degradation rates of mRNA [31, 39, 40] and proteins[41, 42] and transcription factor-chromatin [43-48] and protein-protein interaction rates [49, 50]. In developing animals, complete cell lineage trees can be reconstructed [51]. Such information in principle allows describing gene regulatory networks at full molecular detail [52-55]. Indeed, besides basic network motives [10, 56-60], larger systems have been described at a detailed molecular level, including the pluripotency network [61, 62], the cell cycle [63] and the circadian clock [64]. Since a complete Gillespie simulation [65] of large networks is computationally too time-expensive, networks including microscopic stochastic elements have been treated by kinetic coarse-graining [66] , kinetically separating stochastic and deterministic building blocks [67] or piecewise deterministic Markov processes [68]. However, implementing a simulation environment for a gene regulatory network including molecular details requires complex mathematics and thus is limited to a few specialized groups and network-specific realizations.

Here, we report a computer aided interactive gene network simulation tool (CaiNet), which is generally applicable by users with little mathematical knowledge to model the dynamic behavior of gene regulatory networks at full molecular detail including stochastic kinetics, delays and noise. We validate the similarity in dynamics and noise of CaiNet to full Gillespie simulations using an auto feedback motive. We apply CaiNet to the circadian clock and find that external input modulating

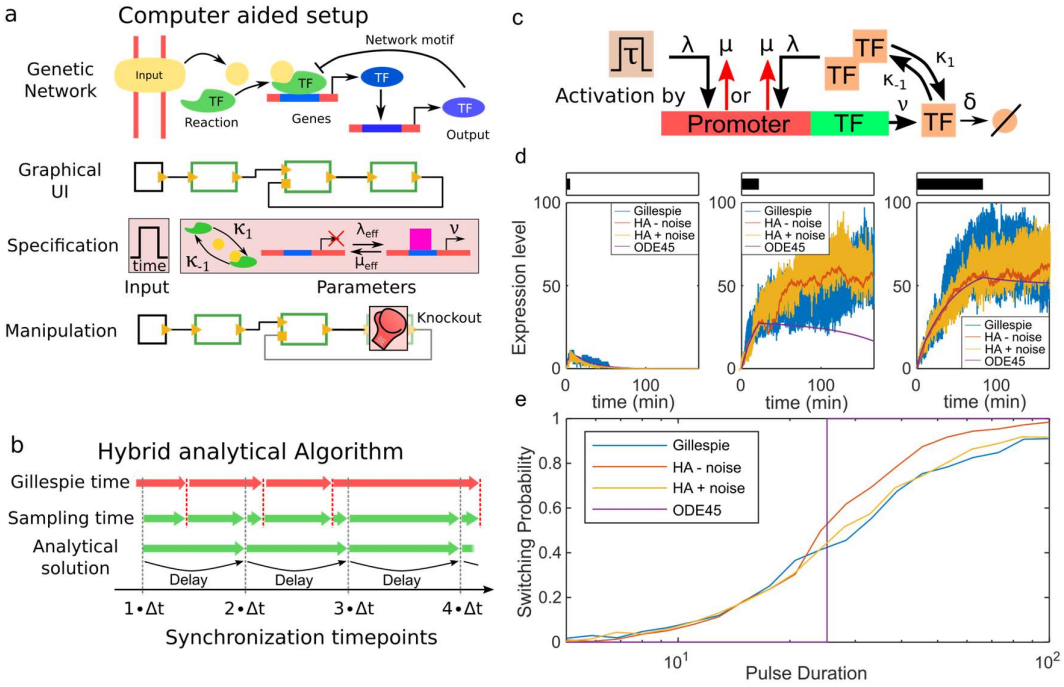


Figure 1: Working principle and validation of CaiNet. (a) Gene Regulatory Networks (GRNs) in CaiNet may comprise extracellular input signals, signal transduction pathways and interactions in a network of genes. These connected building blocks of the regulatory network can be implemented in CaiNet via a GUI, which automatically applies mathematical models for all building blocks of the network. The functions and parameters of these building blocks can be manipulated to emulate experimental conditions such as knockdown experiments. (b) CaiNet uses a half-analytical algorithm to simulate the GRN. Switching between on/off states of genes in simulated with the Gillespie algorithm. The switching trajectory is sampled by analytical solutions of the birth and death processes of the gene products. Chemical reactions such as dimerization and signaling pathways are calculated analytically in time intervals of Δt. At the end of this time interval all calculated numbers of species are synchronized. (c) Network for verification of CaiNet. The network consists of a single transcription factor that forms homodimers to activate its own expression. Such an auto regulatory motif may switch to a high stable expression level. We induce the switching process by applying a pulse of an external activating transcription factor. (d) Time trajectories of the expression level calculated with CaiNet using the Noise add-on and the ODE45 solver of Matlab. (e) Dependency of the switching probability on the pulse duration. Gillespie and CaiNet results overlap well, while the deterministic ODE-approach overlooks stochastic features.

Per/Cry expression levels enables regulating the periodicity of circadian oscillations. We further use CaiNet to simulate the time course of germ line differentiation starting from the pluripotent state.

Results

Functionality and validation of CaiNet

Genetic networks commonly comprise several elements: external signaling inputs, biochemical reactions, signaling motives including several genes and network-specific signaling outputs (Figure 1a, upper panel). We designed a graphical user interface (GUI) for CaiNet to facilitate setting up complex genetic networks (Figure 1a, second panel), inspired by the NetBuilder project [69, 70]. With this GUI, icons representing the network elements 'input', 'biochemical reaction' and 'gene' can be connected intuitively using activating or inhibiting links represented by wires. For each network element, relevant structural and kinetic parameters can be defined (Figure 1a, third panel and Supplementary Table 1). Inputs are associated with a time course, e.g. sinusoidal or rectangular (Supplementary Figure 1). We predefined several biochemical reactions, which include reaction rates such as association or dissociation, and provide analytical solutions to the corresponding ordinary differential equations (ODEs) (Supplementary Figure 2). For genes, the structure of the promoter, e.g. the number of transcription factor binding sites (Supplementary Figure 3), and kinetic parameters such as residence times of transcription factors, transcription and degradation rates can be defined. Genes are described by two states, an on-state that is initiated by association of a transcription factor and leads to production of RNA and an off-state that is assumed upon dissociation of the transcription factor. In addition, we implemented means to manipulate a genetic network by knocking down one or more genes (Figure 1a, lower panel).

To simulate a genetic network using CaiNet, the network laid out in the GUI is automatically transferred to a simulation job based on a hybrid stochastic analytical algorithm (Figure 1b and Methods). Inputs define the initial and boundary conditions of the network. Appropriate analytical solutions are assigned to the biochemical reactions (Supplementary Information Section 1.3). Transitions of genes between on and off states are treated with a Gillespie approach that yields stochastically distributed switching times (Methods). At fixed time points every Δt, analytical and stochastic network elements are synchronized and the momentary simulation output, that is the momentary number of all molecular species including protein and RNA molecules, is tabulated. For synchronization, analytical solutions are directly calculated for each synchronization time point using interval Δt. For stochastic network elements, the Gillespie simulation employs numbers of molecular species determined at the previous synchronization time point. The number of molecular species for the following synchronization time point is obtained by analytical continuation using Equation (8) (Methods). A delay within a stochastic process, e.g. translation of a gene product, can be realized by shifting the synchronization time point of this process with respect to the remaining genetic network (Figure 1b).

The modular design of our hybrid stochastic analytical approach readily enables simulations of increasing complexity (Methods): (i) initial solutions of an ODE solver use average on-times for genes instead of full gene switching kinetics. (ii) ODE solutions can be enhanced by including gene switching kinetics calculated by the Gillepsie approach. (iii) Additional noise in expression levels can be included, drawn from a Poisson distribution.

We validated the hybrid algorithm underlying CaiNet with the network motive of an auto feedback loop that is activated by both a dimer of its gene product and an additional external stimulating input pulse (Figure 1c and Supplementary Table 2). Similar network motifs have been shown to exhibit complex emergent behavior depending on noise [58]. We compared the behavior of this network motive using a full kinetic simulation with the Gillespie algorithm (Supplementary Information Section 2.1.1), simulations using CaiNet in the presence and absence of fluctuations in gene product due to birth and death processes (Methods), and a numerical solution of the corresponding ordinary differential equations (Methods). An immediate advantage of our hybrid stochastic analytical approach compared to Gillespie and ODEs are ca. 100-fold faster simulation times. At a short input pulse of 5 min, the expression level returns to zero for all considered cases (Figure 1d, left panel). At a long input pulse of 100 min, expression levels switch to a high state (Figure 1d, right panel), as expected for the positive auto feedback loop [58]. At intermediate pulse durations, simulations using Gillespie or CaiNet reveal stochastic transitions between the two expression levels, while the ODE solution assumes the outcome predicted by the initial condition. We quantified the switching probability of the auto feedback loop in dependence of the pulse duration by counting the times the expression level increased above a certain threshold in 500 simulation runs for each pulse duration (Figure 1e). As expected, the ODE approach exhibits a binary switching behavior with clearly defined switching pulse duration. In contrast, the Gillespie approach exhibits a sigmoidal switching probability. The switching probabilities of both CaiNet approaches with and without gene product fluctuations are comparable to the Gillespie approach.

Our hybrid algorithm leads to an error in the number of molecular species at the time point of synchronization actually seen by each element that increases with increasing synchronization interval Δt. For the auto feedback loop we observed deviations in the switching probability between CaiNet and the full Gillespie simulation only if Δt exceeded the gene production rate by a factor of 100 (Supplementary Figure 4). The correct outcome of the emergent kinetic feature of the switching probability in the CaiNet simulations indicates that CaiNet well approximates the full Gillespie simulation.

Frequency modulations in the circadian clock

As first application, we used CaiNet to study the temporal behavior of the circadian clock. The circadian clock is a gene network that regulates the 24h day-night rhythm in mammals [71-73]. It consists of an internal incoherent feedback loop composed of the heterodimer Bmal1/Clock that activates expression of Ror and Rev, which in turn activate respectively inhibit expression of Bmal1/Clock (Figure 2a)[73, 74]. In addition, the circadian clock comprises a negative feedback loop in which the Bmal1/Clock-activated heterodimer Per/Cry inhibits its own transcription directly and indirectly via inhibition of Ror and Rev (Figure 2a)[73, 74]. We allowed circadian genes to exhibit low leaky expression. The circadian clock is adjusted by external light and metabolic signals acting on Per/Cry [75, 76] (Figure 2a). Why two functionally similar feedback loops are necessary to drive circadian clock oscillations is not clear.

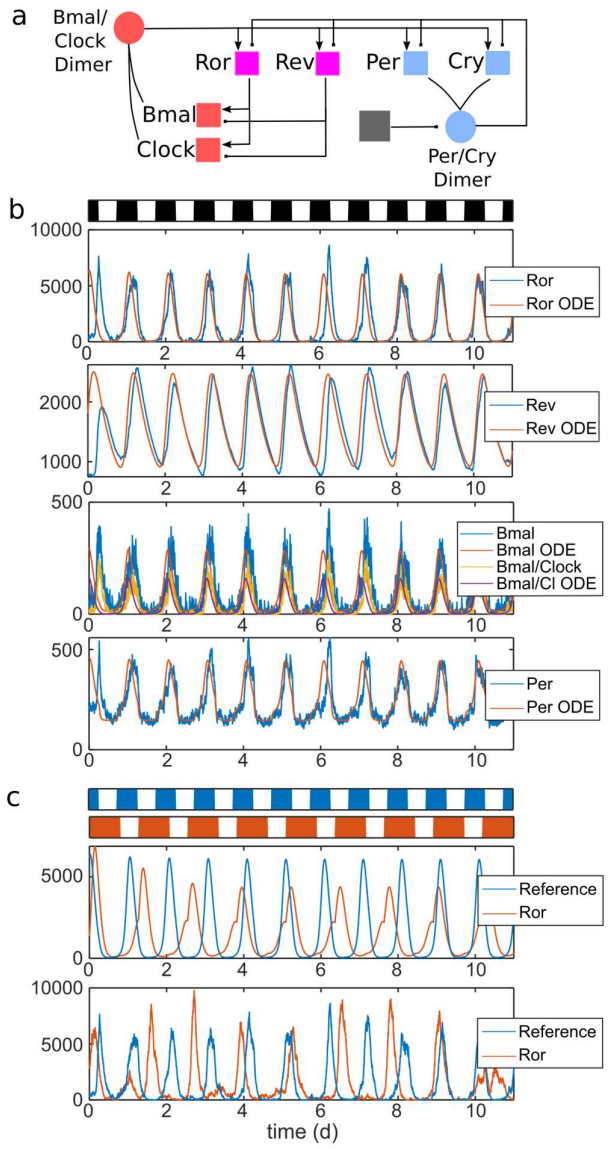


Figure 2: Interplay of two nested loops modulates phase and frequency of circadian oscillations. (a) Simplified layout of the circadian clock network. The inner feedback loop consists of Bmal, clock, Ror and Rev. Ror is the positive and Rev is the negative feedback lobe for Bmal/Clock expression. Per and Cry are an additional negative feedback lobe that can be tuned by external signals. (b) Comparison of protein levels of ODE solutions and solutions including molecular noise. The circadian clock oscillates with a frequency of 24h. The oscillation of the inner loop runs autonomously. Per activity suppresses rising expression levels during an oscillation period. Activating Per every 24h during low expression levels of the inner loop does not impede its oscillations and helps to maintain coherence with a deterministic periodic 24h oscillation. (c) Prolonged Per activity also prolongs down-times of expression levels of Ror, Rev and Bmal/Clock, thereby decreasing the native clock frequency of the circadian clock, even during a period of 10 days. Middle panel: comparison of ODE solutions. Lower panel: solutions including molecular noise.

We first determined whether the internal feedback loop alone allows for circadian oscillations, as previously proposed [77, 78]. We thus implemented the corresponding network in the GUI of CaiNet (Supplementary Figure 5a). We next tested the existence of parameter sets ensuring oscillations by stability analysis of the fix point of the corresponding ODEs (Supplementary Information Section 2.2)[79]. This revealed constraints for kinetic parameters (Equation (1.85) and (1.93) in Supplementary Information) which indeed led to oscillations of the internal incoherent feedback loop when inserted into CaiNet (Supplementary Figure 5b). To restrict the period of oscillations to 24h, we transferred the ODEs to a Liennard equation and determined its eigenfrequency (Supplementary Information equation (1.87)). We thus found constraints for the degradation rates of Ror and Rev that ensured a circadian rhythm of the internal incoherent feedback loop (Supplementary Figure 5b and Supplementary Table 3).

We next added the negative feedback loop of Per/Cry to the internal feedback loop, combined with a periodic input insuring degradation of the Per/Cry dimer (Figure 2a). Again, all components of the circadian clock network exhibited circadian oscillations in their expression levels comparable to experimental oscillations (Figure 2b)[23]. While ODE solutions led to a stable frequency and amplitude of oscillations, the full kinetic description of the network reveals small variations of these parameters, similar to previous observations[64]. By modulating the time course of the input signal, a phase shift in the oscillations of the network can be introduced (Supplementary Figure 6a and b) and the frequency of the oscillations can be altered (Figure 2c). Similarly, an activating input by NFY to Bmal1/Clock [77] introduced a phase shift to the network (Supplementary Figure 7c and d). Our results for the circadian clock network demonstrate that temporally modulated external input signals acting on Per/Cry allow regulating the frequency and phase of circadian oscillations in a nested network topology.

Time course of germ line differentiation

As second application, we used CaiNet to study the temporal behavior of the pluripotency network [6, 80]. In particular, we simulated the time course of differentiation from the naïve state into germ layer cell types upon certain culture conditions [81]. We implemented a network comprised of an input layer including the signaling factors Lif, TGF-b, BMP, Fgf4 and Wnt, signal transduction and feedback layers around the core pluripotency network including Nanog, Sox2 and Oct4 and an output layer of differentiation markers includ-

ing Runx2, Pax6 and Gata6 in the GUI of CaiNet (Figure 3a and Supplementary Information Section 2.3.1)[11, 74, 80]. For the transcription factors Stat3, Sox2, Oct4, Nanog and Klf4 we applied experimentally determined kinetic parameters and promoter structures[42, 82, 83], for the other kinetic parameters of the network elements we chose typical values (Supplementary Table 4).

To test the network for consistency with experiments, we initially traced the time course of expression levels of Klf4, Nanog, Sox2 and Oct4 when changing input conditions sustaining the naïve state, in which Lif and BMP are present, to conditions supporting the primed state, in which TGF-b and Fgf4 are present (Figure 3b). As expected, expression of Klf4 stalled, while expression of the pluripotency factors continued [80]. Next, we used the knockout tool of CaiNet to simulate experi-

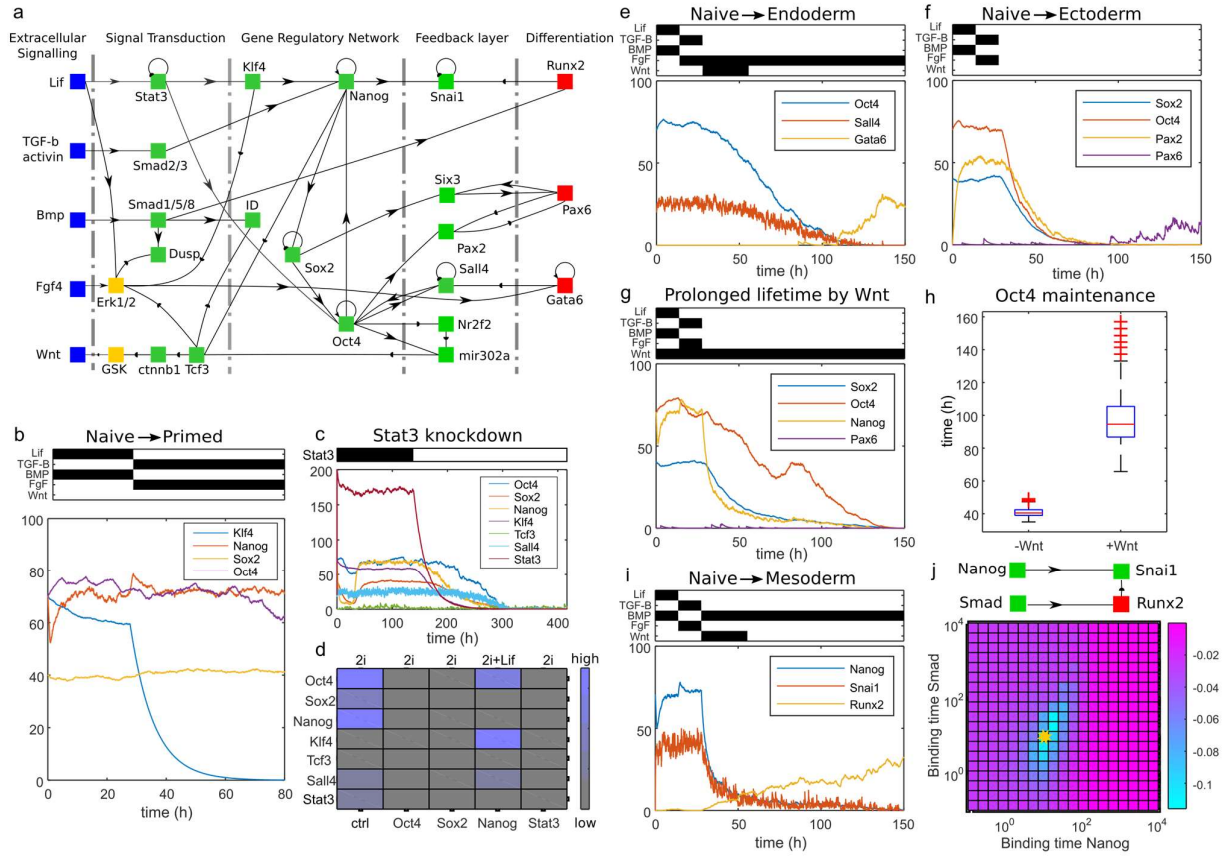


Figure 3: Temporal regulation of Pluripotency and differentiation. (a) Layout of the pluripotency network. The regulatory network is split into different functional layers. The signal transduction layer translates extracellular signals into transcription factors. These interact with core-pluripotency factors and differentiation factors. The feedback layer propagates the state of the pluripotency network to the differentiation factors. (b) Transition from naïve to primed pluripotency. Lif and BMP signaling maintain the activity of Klf4, Nanog, Sox2 and Oct4. When Lif and BMP are replaced by Fgf and Tgf-β, all pluripotency factors except Klf4 maintain their expression levels. (c and d) Simulation of knockdown experiments of core pluripotency factors. In 2i conditions the core pluripotency network is active. While knockdown of all pluripotency factors deactivates the core network, Nanog knockdown in presence of Lif signal shows weak activity of Oct4, Sall4 and Klf4 (e) Transition from naïve pluripotency to the endoderm state. Fgf indirectly activates the endoderm differentiation marker Gata6 which adopts a high expression level since the pluripotency factors Oct4 and Sall4 are inactive (f) Transition from naïve pluripotency to the enctoderm state. Due to the stable feedback between Six3 and Pax6, ectoderm is the default differentiation state if no signal is present. (g) Wnt signaling inhibits the transcription factor Tcf3, which is a down regulator of the core pluripotency network. Thus Wnt also inhibits ectoderm differentiation (h) Wnt signaling determines how low the core network remains active once all activating signals have been shut down (i) Transition from naïve pluripotency to the mesoderm state. BMP activates the mesoderm differentiation maker Runx2 (j) upper panel: network motif for the interpretation of pluripotency and differentiation signals. Lower panel: heatmap of Smad and Nanog binding times revealing optimal values for Runx2 activity.

ments in which Stat3 has been knocked down in 2i culture conditions [11] (Figure 3c and Supplementary Information Section 2.3.4). As expected, expression levels of important signaling factors including Nanog, Sox2 and Oct4 declined within two days. Analogously, knockdown of essential signaling factors resulted in vanishing expression levels of all other factors (Figure 3d). The knockdown of Nanog in presence of 2i and Lif culturing conditions showed low levels of Oct4, Sall4 and Klf4 activity as observed in experiments [11].

Next, we simulated differentiation from the naïve state into the endoderm state by successively altering input conditions within one day from Lif/Bmp to TGF-b/Fgf to Fgf/Wnt, and monitoring the expression levels of Oct4, Sall4 and the differentiation marker Gata6 (Figure 3e) [80, 81, 84]. We found that Sall4, which represses Gata6, is continuously expressed during naïve and primed states until the core pluripotency network including Oct4 is downregulated upon transition to the endoderm line. Conversely, Gata6 was not expressed in naïve state and only showed marginal expression in the primed state due to direct activation by Fgf4, but increased in levels after downregulation of its repressor Sall4 by the core network. Similarly, we simulated differentiation into the ectoderm state (Figure 3f). We observed that differentiation into ectoderm, as indicated by an increase in the marker gene Pax6, occurred as default if no input signals were present after the primed state, consistent with expectations [85]. In contrast, differentiation into ectoderm was prohibited if Wnt signaling was present [86] (Figure 3g). Our simulation revealed that this is attributed to a prolonged presence of Oct4 when negative feedback loops activated by Tcf3 are kept inactive (Figure 3h). A similar finding was suggested previously [87]. Finally, when we altered the inputs to Bmp only after the primed state, differentiation occurred into the mesoderm state, indicated by an increase of the marker gene Runx2 (Figure 3i).

In our layout of the pluripotency network, differentiation into the germ layer typically occurs via a regulatory motive in which a pluripotency factor inhibits and an input signal activates a marker gene. Similar regulatory motives are also found in other complex network structures [88]. We tested how the chromatin interaction times of the two adversely acting transcription factors affected transcription of the marker gene Runx2 in the subnetwork including Nanog and Smad1/5/8 (Figure 3j, upper panel). We found a strong dependency of Runx2 activity on the chromatin residence times of Nanog and Smad1/5/8 (Figure 3j, lower panel and Supplementary Information Section 2.3.5).

Discussion

CaiNet is designed to assist users in setting up and simulating complex gene regulatory networks at molecular detail. We included analytical solutions to common biochemical reactions in the hybrid stochastic-deterministic simulation approach, resulting in superior simulation speed compared to Gillespie or ODE-based approaches. This library can be extended in the future, as more analytical solutions become available. Novel networks are automatically compiled into the simulation algorithm. This allows to quickly assessing the effect of changes in network structure on the behaviour of the network, without cumbersome recalculation of the network. The knockdown tool, which either reduces or abolishes expression of a gene, further facilitates validation of the network structure through comparison with experiments or by evoking testable predictions.

Our modular algorithm enables adding more and more realistic kinetic behaviour and molecule noise to initial ODE solutions. For the pulse-triggered auto feedback motive, noise led to transitions between stable states[58] in contrast to a sharp threshold behaviour observed in the ODE solution. For the circadian clock, the modular approach enabled us to find parameters[79] allowing for oscillations of network components. When simulating the pluripotency network including molecular noise, we observed a decay of Oct4 abundance with decay time depending on the input condition, revealing a memory inherent in the network structure as previously suggested [89]. Furthermore, our simulations revealed a dependency of network performance on the kinetics of gene switching [61], respectively the kinetics of transcription factor-chromatin interactions. Overall, the influence of molecular kinetics and proper noise on the performance of gene regulatory networks[52, 55] can be readily appreciated using CaiNet.

In setting up the gene regulatory network for the circadian clock and the pluripotency network, we conceded several simplifications to an educational demonstration of CaiNet. First, our choice of participating genes compromises between comprehensiveness and consistency. The layout of the networks can be improved in the future as more information becomes available. Second, we simplified regulatory mechanisms. In particular, the relationship between binding of a transcription factor and activation of a gene neglects for example the activity of cofactors or epigenetic alterations of the gene locus. Moreover, we neglected regulatory elements such as posttranslational modifications or the

presence of RNA binding proteins [5]. Since a series of stochastic processes results in a single effective stochastic process, missing mechanisms can be supplemented by appropriate delays. Thus, while future additions of a more complex regulation of gene activity or further regulatory elements might result in different temporal network behaviour, however, additional feedback motives are expected to impact network behaviour most.

In principle, smart application of the features already available in CaiNet might enable implementing many missing regulatory elements. Posttranslational modifications to a protein for example could be achieved by directly linking two genes and proper assignment of parameters. Reaction diffusion processes might be realized in a similar fashion. On the other hand, CaiNet meets limitations when applied to networks of different type, for example networks of fast biochemical reactions that might conflict with temporal separation in the hybrid algorithm. For gene regulatory networks, CaiNet represents a readily applicable and versatile simulation environment to unravel kinetic network properties at molecular detail.

Methods

We develop a hybrid algorithm that is composed of stochastic and deterministic components. Stochastic components include common promoter structures (see Supplementary Information Section 1.2) and are simulated with the Gillespie-Algorithm. The deterministic components include common signaling pathway motifs and chemical reactions (see Supplementary Information Section 1.3) and are calculated with predefined analytical solutions. We decouple all components during the simulation to increase computational speed. We let the blocks communicate their current molecule number on defined synchronization time points. The synchronization time points allow us to include constant delays without computational effort. Analytical solutions for gene product synthesis allow for efficient simulation of initiation and termination events in transcription and translation. In addition, we allow external user-defined inputs that can change over time.

Pseudo Code:

for *(t=0 ; t<T ; t=t+Δt)*

Apply external inputs to *n(t)*

Apply constant Delay by fetching numbers from the past *n(t)=n(t-delay)*

for every *block* **in** *Stochastic-blocks*

n(t) =*f*(n(t- Δt),block)

// Methods and Supplementary Information Section 1.2

endfor

for every *block* **in** *Deterministic-blocks*

n(t) =*g*(n(t- Δt),block)

// Supplementary Information 1.3

endfor

t=t+dt

endfor

Stochastic building block of gene switching

We model genes as two-state systems [90] that produce a gene product while in the on-state (birth process) and are silent in the off-state (Supplementary Figure 3a). Proteins degrade in both states of the gene (death process). The stochastic rates for switching to the on-state is called by $\lambda(\vec{n})$ and the rates for switching to the off state are called $\mu(\vec{n})$. We assume that the switching events are Poisson processes. As a result, the time of a stochastic event is determined by the random number X and the current switching rate constant

$$t_i = \frac{-\log(X)}{\alpha} \qquad \text{where} \quad \alpha = \begin{cases} \lambda, \text{ Gene off} \\ \mu, \text{ Gene on} \end{cases} \tag{1}$$

Using (1) we obtain a time series of stochastic switching events $\tau = \{t_1, t_2, ...\}$ for each gene that determine the time evolution of the gene product during the time interval Δt between two synchronization time points. From this time-series we construct the function $a(t)$ that describes the time-trajectory of the promoter states. We use this trajectory to calculate the expression level of the gene.

Gene expression including delays

In the on-state, the gene produces products with the rate ν. The products are degraded. We assume that degradation is a Poisson process with rate δ for each

product. Each individual gene has its own birth and death rates ν, δ. The differential equation for gene product synthesis and degradation is given by

$$\dot{n} = a(t)\nu - \delta n \tag{2}$$

where we introduced the time dependent state $a(t)$ of the gene that adopts the values zero and one. The rate constant ν summarizes the many steps included in the synthesis of a gene product. In case of protein synthesis these steps include initiation of transcription, elongation of RNA, termination of transcription, initiation of translation, elongation of the protein and termination of translation. In addition, the mRNA needs time to be transported to ribosomes and might be subject to posttranslational modifications. We assume that each of the aforementioned processes has a single rate-limiting step as it was shown for splicing [91]. Rate limiting steps change the time trajectory of the expression level n by introducing delays with rate constants $\beta_1 ... \beta_N$. In particular, they may smooth bursts of gene expression. To determine the resulting function $n(t)$ we solve the general system of ODEs for an arbitrary number of delay processes. Using the switching function $a(t)$ for a single gene we arrive at the system of differential equations

$$\begin{aligned} \dot{c}_1 &= a(t)\nu - \beta_1 c_1 \\ \dot{c}_2 &= \beta_1 c_1 - \beta_2 c_2 \\ &... \\ \dot{c}_N &= \beta_{N-1} c_{N-1} - \beta_N c_N \end{aligned} \tag{3}$$

where $c_N \triangleq n$ represents the output expression level of the gene. Our first step to solve the system for $c_N(t)$ is to calculate the Laplace transform of all equations in (3). The result for the first and the i-th equation is

$$\begin{aligned} \tilde{c}_1 &= \frac{1}{p+\beta_1}\nu\tilde{a} \\ \tilde{c}_i &= \frac{\beta_{i-1}}{p+\beta_i}\tilde{c}_{i-1} \end{aligned} \tag{4}$$

where p the transform variable and $\tilde{a}$ is the Laplace transform of a. With these results we can easily find an equation for the Laplace-transform of the final product c_N

$$\tilde{c}_N = \frac{\prod_{i=1..N-1}\beta_i}{\prod_{i=1..N}(p+\beta_i)}\nu\tilde{a} \tag{5}$$

Before we calculate the inverse Laplace transform to obtain $c_N(t)$ we simplify the denominators in (5) by partial fraction composition with the coefficients α_i

$$\begin{aligned} &\frac{\prod_{i=1..N-1}\beta_i}{\prod_{i=1..N}(p+\beta_i)} = \sum_i \frac{\alpha_i}{p+\beta_i} \\ &\text{where} \quad \alpha_i = \frac{\prod_{n=1..N-1}\beta_n}{\prod_{n\neq i}(-\beta_i+\beta_n)} \end{aligned} \tag{6}$$

For few delays this simplification can most easily be verified by plugging in α_i and rewriting the right hand side of (6). For a high number of delays the result can be derived by the well-known technique of partial fraction decomposition. We plug this result in (5) and obtain

$$\tilde{c}_N = \nu\tilde{a}\sum_i \frac{\alpha_i}{p+\beta_i} \tag{7}$$

We next use the multiplication theorem for the Laplace transform and find

$$c_N = \sum_i \alpha_i \nu \int \exp(-\beta_i(t-t'))\cdot a(t')\,\mathrm{d}t' \tag{8}$$

Since $a(t)$ only adopts the constant values zero or one this integral can be solved analytically.

Besides single time-limiting steps, elongation of transcripts or Amino Acid chains can introduce a delay in terms of a well-defined time-span in contrast to an exponential distribution. We add these deterministic delays directly in the algorithm by referring to previous synchronization time points. A limitation of this procedure is that the delays can only be given as multiples of the synchronisation time-step Δt.

The above formalism does not limit the algorithm to models of exponential degradation. In (Supplementary Information Section 1.3.2) we demonstrate that the degradation model can be modified by adding an additional equation to the system (3).

Birth and death noise

We are able to emulate a stochastic Gillespie simulation of stochastic birth and death processes by calculating the mRNA produced during a burst of transcription according to the Poisson distribution. Assuming a production rate of ν and an on-time of Δt the probability

distribution p(k,N) for k produced molecules with the expectation value $\nu\Delta t$ is

$$p(k,N)=\frac{N^k}{k!}\exp(-N) \quad (9)$$

In our simulator we draw a random number from the Poisson distribution and obtain the amount of gene products added. The magnitude of the variance in relation to the expectation value is $1/\sqrt{N}$ indicating that this emulation of Birth and death noise is most relevant for small molecule numbers.

Since we apply the noise in the first step of gene product synthesis, that method presented here reproduces correct expectation value and variance for arbitrary gene-switching and delay rate constants-

Top-Down design of noisy networks in CaiNet

Networks assembled in CaiNet can be investigated in a deterministic mode with gene-switching noise and birth and death noise switched off. For each individual gene that is simulated n this deterministic mode, we omit the random number generation in (1) and replace the production rate ν by the average production rate considering on/off times of the gene

$$\nu_{\text{Deterministic}}=\nu\frac{\lambda}{\lambda+\mu}=p_{on}\nu \quad (10)$$

In addition, birth and death noise of gene products (Methods Section 1.1.2) and noise in chemical reactions (Supplementary Information 1.3.6) can be turned on and off for individual species. Using this feature the influence of noise of single processes or species on the network can be determined.

Acknowledgements

We thank David Suter (EPFL, Lausanne, Switzerland) for comments on the pluripotency network. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme [637987 ChromArch to J.C.M.G.].

Author contributions

J.H. and J.C.M.G. designed the study; J.H. performed calculations and programmed CaiNet; J.H. and J.C.M.G. wrote the manuscript.

Code availability

The CaiNet software as well as implementations of the circadian clock and pluripotency networks will be freely available after publication.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Schmidt-Heck, W. et al. Fuzzy modeling reveals a dynamic self-sustaining network of the GLI transcription factors controlling important metabolic regulators in adult mouse hepatocytes. *Molecular Biosystems* **11**, 2190-2197 (2015).
2. Edwards, J.S., Ibarra, R.U. & Palsson, B.O. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology* **19**, 125-130 (2001).
3. Orlando, D.A. et al. Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature* **453**, 944-U978 (2008).
4. Dibner, C., Schibler, U. & Albrecht, U. The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks. *Annual Review of Physiology* **72**, 517-549 (2010).
5. Li, M. & Belmonte, J.C.I. Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics* **18**, 180-191 (2017).
6. Chickarmane, V., Troein, C., Nuber, U.A., Sauro, H.M. & Peterson, C. Transcriptional dynamics of the embryonic stem cell switch. *Plos Computational Biology* **2**, 1080-1092 (2006).
7. Pfeuty, B., Kress, C. & Pain, B. Network Features and Dynamical Landscape of Naive and Primed Pluripotency. *Biophysical Journal* **114**, 237-248 (2018).
8. Davidson, E.H. et al. A genomic regulatory network for development. *Science* **295**, 1669-1678 (2002).
9. Jacob, F. & Monod, J. GENETIC REGULATORY MECHANISMS IN SYNTHESIS OF PROTEINS. *Journal of Molecular Biology* **3**, 318-& (1961).
10. Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**, 450-461 (2007).
11. Dunn, S.J., Martello, G., Yordanov, B., Emmott, S. & Smith, A.G. Defining an essential transcription factor program for naive pluripotency. *Science* **344**, 1156-1160 (2014).
12. Xu, H.L., Ang, Y.S., Sevilla, A., Lemischka, I.R. & Ma'ayan, A. Construction and Validation of a Regulatory Network for Pluripotency and Self-Renewal of Mouse Embryonic Stem Cells. *Plos Computational Biology* **10** (2014).
13. Lee, T.I. et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799-804 (2002).
14. Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14796-14799 (2003).
15. Shmulevich, I., Dougherty, E.R., Kim, S. & Zhang, W. Probabilistic Boolean networks: a rule-based

uncertainty model for gene regulatory networks. *Bioinformatics* **18**, 261-274 (2002).

16. Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using Bayesian networks to analyze expression data. *Journal of Computational Biology* **7**, 601-620 (2000).
17. Nachman, I., Regev, A. & Friedman, N. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* **20**, 248-256 (2004).
18. Bonnaffoux, A. et al. WASABI: a dynamic iterative framework for gene regulatory network inference. *Bmc Bioinformatics* **20** (2019).
19. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).
20. Blake, W.J., Kaern, M., Cantor, C.R. & Collins, J.J. Noise in eukaryotic gene expression. *Nature* **422**, 633-637 (2003).
21. Raser, J.M. & O'Shea, E.K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-1814 (2004).
22. Krebs, A.R. et al. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell* **67**, 411-+ (2017).
23. Suter, D.M. et al. Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* **332**, 472-474 (2011).
24. Larsson, A.J.M. et al. Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251-+ (2019).
25. Ochiai, H. et al. Genome-wide analysis of transcriptional bursting-induced noise in mammalian cells. *bioRxiv*, 736207 (2019).
26. Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A. & Singer, R.H. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. *Science* **332**, 475-478 (2011).
27. Fuchs, G. et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biology* **15** (2014).
28. Righini, M. et al. Full molecular trajectories of RNA polymerase at single base-pair resolution. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 1286-1291 (2018).
29. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *Rna-a Publication of the Rna Society* **14**, 1959-1972 (2008).
30. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
31. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnology* **29**, 436-U237 (2011).
32. Li, G.W., Burkhardt, D., Gross, C. & Weissman, J.S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157**, 624-635 (2014).
33. Weinberg, D.E. et al. Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Reports* **14**, 1787-1799 (2016).
34. Moffitt, J.R. & Zhuang, X. RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). *Visualizing Rna Dynamics in the Cell* **572**, 1-49 (2016).
35. Beck, M. et al. The quantitative proteome of a human cell line. *Molecular Systems Biology* **7**, 549 (2011).
36. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology* **7**, 548 (2011).
37. Strebinger, D. et al. Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions. *Molecular Systems Biology* **15**, e9002 (2019).
38. Bothma, J.P., Norstad, M.R., Alamos, S. & Garcia, H.G. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell* **173**, 1810-+ (2018).
39. Tani, H. et al. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Research* **22**, 947-956 (2012).
40. Blumberg, A. et al. Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. *BioRxiv*, 690644 (2019).
41. Knop, M. & Edgar, B.A. Tracking protein turnover and degradation by microscopy: photo-switchable versus time-encoded fluorescent proteins. *Open Biology* **4** (2014).
42. Alber, A.B., Paquet, E.R., Biserni, M., Naef, F. & Suter, D.M. Single Live Cell Monitoring of Protein Turnover Reveals Intercellular Variability and Cell-Cycle Dependence of Degradation Rates. *Molecular Cell* **71**, 1079-+ (2018).
43. Mazza, D., Abernathy, A., Golob, N., Morisaki, T. & McNally, J.G. A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Research* **40**, e119 (2012).
44. Gebhardt, J.C.M. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature Methods* **10**, 421-+ (2013).
45. Ho, H.N., Zalami, D., Kohler, J., van Oijen, A.M. & Ghodke, H. Identification of Multiple Kinetic Populations of DNA-Binding Proteins in Live Cells. *Biophysical Journal* **117**, 950-961 (2019).
46. Reisser, M., Hettich, J., Kuhn, T. & Gebhardt, J.C.M. Inferring quantity and qualities of superimposed reaction rates in single molecule survival time distributions. *bioRxiv*, 679258 (2019).
47. Azpeitia, E. & Wagner, A. Short residence times of DNA-bound transcription factors can reduce gene expression noise and increase the transmission of information in a gene regulation system. *bioRxiv*, 776955 (2019).
48. Nutiu, R. et al. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnology* **29**, 659-U146 (2011).
49. Luker, K.E. et al. Kinetics of regulated protein-protein interactions revealed with firefly luciferase

complementation imaging in cells and living animals. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12288-12293 (2004).

50. Brezovich, A., Schuschnig, M., Ammerer, G. & Kraft, C. An in vivo detection system for transient and low-abundant protein interactions and their kinetics in budding yeast. *Yeast* **32**, 355-365 (2015).
51. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nature Biotechnology* **36**, 469-+ (2018).
52. Kerszberg, M. Noise, delays, robustness, canalization and all that. *Current Opinion in Genetics & Development* **14**, 440-445 (2004).
53. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* **9**, 770-780 (2008).
54. Swift, J. & Coruzzi, G.M. A matter of time - How transient transcription factor interactions create dynamic gene regulatory networks. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1860**, 75-83 (2017).
55. Kaern, M., Elston, T.C., Blake, W.J. & Collins, J.J. Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics* **6**, 451-464 (2005).
56. Hasty, J., Pradines, J., Dolnik, M. & Collins, J.J. Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 2075-2080 (2000).
57. Lewis, J. Autoinhibition with transcriptional delay: A simple mechanism for the zebrafish somitogenesis oscillator. *Current Biology* **13**, 1398-1408 (2003).
58. Chaudhury, S. Modeling the effect of transcriptional noise on switching in gene networks in a genetic bistable switch. *Journal of Biological Physics* **41**, 235-246 (2015).
59. Kepler, T.B. & Elston, T.C. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophysical Journal* **81**, 3116-3136 (2001).
60. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593 (2000).
61. Lin, Y.T., Hufton, P.G., Lee, E.J. & Potoyan, D.A. A stochastic and dynamical view of pluripotency in mouse embryonic stem cells. *Plos Computational Biology* **14**, 24 (2018).
62. Chickarmane, V., Olariu, V. & Peterson, C. Probing the role of stochasticity in a model of the embryonic stem cell - heterogeneous gene expression and reprogramming efficiency. *Bmc Systems Biology* **6**, 12 (2012).
63. Barik, D., Ball, D.A., Peccoud, J. & Tyson, J.J. A Stochastic Model of the Yeast Cell Cycle Reveals Roles for Feedback Regulation in Limiting Cellular Variability. *Plos Computational Biology* **12** (2016).
64. Forger, D.B. & Peskin, C.S. Stochastic simulation of the mammalian circadian clock. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 321-324 (2005).
65. Gillespie, D.T. EXACT STOCHASTIC SIMULATION OF COUPLED CHEMICAL-REACTIONS. *Journal of Physical Chemistry* **81**, 2340-2361 (1977).
66. Sinitsyn, N.A., Hengartner, N. & Nemenman, I. Adiabatic coarse-graining and simulations of stochastic biochemical networks. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10546-10551 (2009).
67. Lecca, P., Bagagiolo, F. & Scarpa, M. Hybrid deterministic/stochastic simulation of complex biochemical systems. *Molecular Biosystems* **13**, 2672-2686 (2017).
68. Lin, Y.T. & Buchler, N.E. Efficient analysis of stochastic gene dynamics in the non-adiabatic regime using piecewise deterministic Markov processes. *Journal of the Royal Society Interface* **15** (2018).
69. Brown, C.T. et al. New computational approaches for analysis of cis-regulatory networks. *Developmental Biology* **246**, 86-102 (2002).
70. Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20**, 3289-3291 (2004).
71. Feldman, J.F. GENETIC APPROACHES TO CIRCADIAN CLOCKS. *Annual Review of Plant Physiology and Plant Molecular Biology* **33**, 583-608 (1982).
72. Asgari-Targhi, A. & Klerman, E.B. Mathematical modeling of circadian rhythms. *Wiley Interdisciplinary Reviews-Systems Biology and Medicine* **11**, 19 (2019).
73. Takahashi, J.S. Transcriptional architecture of the mammalian circadian clock. *Nature Reviews Genetics* **18**, 164-179 (2017).
74. Ogata, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **27**, 29-34 (1999).
75. Pilorz, V. et al. A Novel Mechanism Controlling Resetting Speed of the Circadian Clock to Environmental Stimuli. *Current Biology* **24**, 766-773 (2014).
76. Camacho, F. et al. Human casein kinase I delta phosphorylation of human circadian clock proteins period 1 and 2. *Febs Letters* **489**, 159-165 (2001).
77. Xiao, J. et al. Transcription Factor NF-Y Is a Functional Regulator of the Transcription of Core Clock Gene Bmal1. *Journal of Biological Chemistry* **288**, 31930-31936 (2013).
78. Relogio, A. et al. Tuning the Mammalian Circadian Clock: Robust Synergy of Two Loops. *Plos Computational Biology* **7**, 18 (2011).
79. Efimov, D.V. & Fradkov, A.L. Yakubovich's oscillatority of circadian oscillations models. *Mathematical Biosciences* **216**, 187-191 (2008).
80. Morgani, S., Nichols, J. & Hadjantonakis, A.K. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. *Bmc Developmental Biology* **17** (2017).
81. Loh, K.M. et al. Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451-467 (2016).

82. Chen, J.J. et al. Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell* **156**, 1274-1285 (2014).
83. Xie, L.Q. et al. A dynamic interplay of enhancer elements regulates Klf4 expression in naive pluripotency. *Genes & Development* **31**, 1795-1808 (2017).
84. Meng, Y. et al. GATA6 phosphorylation by Erk1/2 propels exit from pluripotency and commitment to primitive endoderm. *Developmental Biology* **436**, 55-65 (2018).
85. Munoz-Sanjuan, I. & Brivanlou, A.H. Neural induction, the default model and embryonic stem cells. *Nature Reviews Neuroscience* **3**, 271-280 (2002).
86. Osteil, P. et al. Dynamics of Wnt activity on the acquisition of ectoderm potency in epiblast stem cells. *Development* **146**, 14 (2019).
87. Papatsenko, D. et al. Single-Cell Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal. *Stem Cell Reports* **5**, 207-220 (2015).
88. Brandman, O., Ferrett, J.E., Li, R. & Meyer, T. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* **310**, 496-498 (2005).
89. Papatsenko, D., Waghray, A. & Lemischka, I.R. Feedback control of pluripotency in embryonic stem cells: Signaling, transcription and epigenetics. *Stem Cell Research* **29**, 180-188 (2018).
90. Peccoud, J. & Ycart, B. MARKOVIAN MODELING OF GENE-PRODUCT SYNTHESIS. *Theoretical Population Biology* **48**, 222-234 (1995).
91. Audibert, A., Weil, D. & Dautry, F. In vivo kinetics of mRNA splicing and transport in mammalian cells. *Molecular and Cellular Biology* **22**, 6706-6718 (2002).

Supplementary Information

Computer aided interactive gene network simulations including stochastic molecular kinetics and noise

Johannes Hettich and J. Christof M. Gebhardt[#]
Institute of Biophysics, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm
[#]To whom correspondence should be addressed: christof.gebhardt@uni-ulm.de

1 Elements of CaiNet

1.1 Inputs

We designed input-elements that feed a user-defined time trajectory into the network. Such elements may be used to simulate external stimuli or noise-less constant expression levels. Figure 1 shows example input trajectories.

1.2 Effective on/off rates of the stochastic two state model

To simplify our computational approach, we assume that the complex dynamics on a promoter can be reduced to a two state model. For each promoter structure below we calculate effective on- and off-rates λ_{eff} and μ_{eff} that correspond to a two state model.

1.2.1 Activation by a single TF

For the most simplified case of activation of a gene we assume that a promoter is activated once a TF binds (Figure 3a). This means that the on-rate is given by the arrival rate of the TF at the promoter. Once a TF has arrived at the Promoter the gene is 'on' for the time $1/\mu$. This time can vary depending on the TF and the promoter of the gene. In not otherwise stated we assumed that the binding time of the TF corresponds to this on-time.

$$\begin{aligned}\lambda_{eff} &= n_{TF}\lambda_0 \\ \mu_{eff} &= \mu\end{aligned} \tag{1.1}$$

where λ_0 is the arrival rate of a single transcription factor.

1.2.2 Promoter with AND-logic

The AND-logic refers to a promoter that is only activated if TF_1 and TF_2 up to TF_N are bound (Figure 3b). Once a single TF leaves, the activation criterion is immediately violated and the promoter is off. Therefore, the off-rate of the promoter μ_{eff} is the sum over all off-rates of the TFs μ_{TFi}.

$$\mu_{eff} = \sum_{i=1}^{N} \mu_{TFi} \tag{1.2}$$

Combinatorically, we can also write down the probability of the promoter to be on as the product of the probability of all TFs to be bound

$$p_{on} = \prod_i p_{on,TFi} = \prod_i \lambda_{TFi}/(\lambda_{TFi}+\mu_{TFi}) \tag{1.3}$$

where $\lambda_{TFi} = \lambda_{TFi,0} n_{TF,i}$ is the arrival rate of a TF at the promoter. From $p_{on,eff}$ and μ_{eff} we can calculate the on-rate of the promoter

$$\lambda_{eff} = p_{on}/(1-p_{on})\mu_{eff} = K\mu_{eff} \tag{1.4}$$

1.2.3 Promoter with OR logic

When a promoter is active if TF1 or TF2 up to TFN or a combination of all is bound, we refer to this promoter as OR-logic (Figure 3c). We start the calculation of effective rates with the on-rates of individual TFs λ_{TFi}. Since the arrival of any TF is enough to activate the promoter, the on-rate is

$$\lambda_{eff} = \sum_i \lambda_{TFi} \tag{1.5}$$

The probability to be on can be calculated combinatorically with $p_{on,TFi}$ implicitly defined in (1.3)

$$p_{on} = 1-(1-p_{on,TF1})(1-p_{on,TF2})...(1-p_{on,TFN}) \tag{1.6}$$

With (1.4) we find the effective off rate of the promoter

$$\mu_{eff} = \lambda_{eff}(1-p_{on})/p_{on} \tag{1.7}$$

1.2.4 Promoter with cooperating TFs

We next describe a promoter where TF1 enhances the binding time of TF2 (Figure 3d). We first calculate the on- and off-rate of both TFs being bound to the promoter.

$$\begin{aligned} \lambda_{eff,TF1,TF2} &= \frac{(\mu_1+\mu_2)p_{on}}{1-p_{on}} \\ \mu_{eff,TF1,TF2} &= \mu_1+\mu_2 \end{aligned} \tag{1.8}$$

We add this effective on and off rates to the or-promoter described above. These effective rates enter equations (1.5) and (1.7) to calculate the effective on/off rates of the gene.

1.2.5 Competitive Repression

We now enhance all promoters developed above by an additional feature, which is the blocking of the promoter by a repressor. We assume that the promoter cannot be activated once a repressor is bound. Once a gene is activated however, the repressor does not abort expression. Therefore, the repressor directly affects the effective on-rate of an arbitrary promoter while the off-rate remains unchanged

$$\begin{aligned} \lambda_{eff,repressor} &= \lambda_{eff} \cdot p_{off,\mathrm{Repressor}} \\ \mu_{eff,repressor} &= \mu_{eff} \end{aligned} \tag{1.9}$$

To proof this, we write down the differential equations describing the change in the blocked and free promoter populations. We denote blocked promoters with b free but inactive promoters with f and active promoters with a

$$\begin{aligned} \dot{b} &= -\mu_R b + \lambda_R f \\ \dot{f} &= \mu_R b + \mu_{eff} a - \lambda_{eff} f - \lambda_R f \end{aligned} \tag{1.10}$$

Assuming that the changes in repressor-population are small, we find

$$b = \frac{\lambda_R}{\mu_R} f = K_R f \tag{1.11}$$

If we now calculate the sum of blocked and unblocked promoters, we obtain

$$p = b + f = (1+K_R)f \tag{1.12}$$

We add up equations (1.10) plug in (1.12) and obtain

$$\dot{p} = \mu a - \frac{\lambda_{eff}}{1+K_R} p \tag{1.13}$$

From this result we conclude for the new effective rate modified by the repressor

$$\lambda_{eff,repressor} = \lambda_{eff}\left(1+K_R\right)^{-1} \tag{1.14}$$

1.2.6 Noncompetitive repression

For a non-competitive repressor the gene is off if the repressor is bound. Therefore, the new probability to be active is

$$p_{on} = p_{on,activatingTF}\, p_{off,repressor} \tag{1.15}$$

To obtain the effective rates in the two-state model, we assume that binding of the repressor during the on-time of the gene is unlikely. With this assumption we can calculate the new on-rate with

$$\lambda_{eff,repressor} = \mu_{eff}\frac{p_{on}}{1-p_{on}} \tag{1.16}$$

1.3 Chemical reactions

Biochemical reactions may dramatically slow down pure Gillespie approaches due to high protein numbers coming with a high frequency of reaction events. At the same time, hybrid approaches using ODE-solvers for biochemical reactions can also be slowed down due to computational overhead in switching between algorithms and small integration step sizes due to nonlinearities in the differential equations. We here provide a library of analytical solutions for frequently occurring biochemical reaction motifs in presence of birth and death processes to speed up simulations in the hybrid stochastic deterministic framework. Table 1 and Figure 2 give an overview of biochemical reaction motifs and occurring kinetic variables.

1.3.1 Dimerization

We solve the differential equation of homo- and hetero-dimerization (Figure 2a). We first consider the heterodimerization where species A and B form a heterodimer AB with the dimerization rate λ. The heterodimer AB can degrade into the monomers with rate μ.

$$A + B \underset{\mu}{\overset{\lambda}{\rightleftharpoons}} AB \tag{1.17}$$

We assume that both monomer populations are subject to slow birth and death processes with rates ν_A, ν_B and $\delta_A m_A, \delta_B m_B$ respectively where we introduced the number of monomers m_A, m_B of species A and B. We obtain the system of differential equations

$$\begin{aligned}
\dot{m}_A &= -\lambda m_A m_B + \mu d + \nu_A - \delta_A m_A + \delta_B d \\
\dot{m}_B &= -\lambda m_A m_B + \mu d + \nu_B - \delta_B m_B + \delta_A d \\
\dot{d} &= \lambda m_A m_B - \mu d - \delta_A d - \delta_B d
\end{aligned} \tag{1.18}$$

where the number of heterodimers is abbreviated with d. Adding up the above equation yields the particle conservation law

$$\begin{aligned}
m_A + d &= N_A \\
m_B + d &= N_B
\end{aligned} \tag{1.19}$$

The total numbers of molecules N_A and N_B are subject to small changes in population due to birth and death processes. We plug (1.19) in (1.18) and obtain

$$\frac{1}{\lambda}\dot{d} = -\frac{d}{K} + (N_A - d)(N_B - d) \qquad \text{K=}\frac{\lambda}{\mu + \delta_A + \delta_B} \tag{1.20}$$

The fixed point of the differential equation is obtained by setting the left hand side of (1.20) to zero and solving for d. The two solutions of the quadratic equation are

$$d_2 = \frac{1}{K}\bar{d}(KN_A, KN_B) \qquad d_1 = \frac{N_A N_B}{d_2} \tag{1.21}$$

where we introduced the function

$$\bar{d}(u,v) = \frac{u + v + 1 - \sqrt{(u-v)^2 + 2(u+v) + 1}}{2} \tag{1.22}$$

We use this result to rewrite (1.20) and obtain

$$\frac{\dot{d}}{(d_1 - d)(d_2 - d)} = \lambda \tag{1.23}$$

The solution of this differential equation is given by

$$d = d_2 - \frac{\tilde{\lambda}}{\lambda} \cdot \frac{u_0 \exp(-\tilde{\lambda}t)}{1 - u_0 \exp(-\tilde{\lambda}t)} \qquad \tilde{\lambda} = \lambda(d_1 - d_2) \tag{1.24}$$

where u_0 is given by the initial number of dimers d_0

$$u_0 = \frac{d_2 - d_0}{d_1 - d_0} \tag{1.25}$$

We can apply this solution to the case of a homodimer by substituting

$$\begin{aligned} \lambda' &= 4\lambda \\ K &= \lambda' / (\mu + 2\delta) \\ N_A &= M/2 \\ N_B &= (M-1)/2 \end{aligned} \tag{1.26}$$

with this substitution we obtain the solution for the homo-dimers from (1.24)

$$d = d_2 - \frac{\tilde{\lambda}}{\lambda'} \cdot \frac{u_0 \exp(-\tilde{\lambda}t)}{1 - u_0 \exp(-\tilde{\lambda}t)} \qquad \tilde{\lambda} = \lambda'(d_1 - d_2) \tag{1.27}$$

1.3.2 Targeted degradation

We consider a protein A of amount n that is degraded by a Michaelis-Menten type reaction with enzyme B of amount m (Supplementary Figure 2 panel b). The corresponding system of differential equations is

$$\begin{aligned}\dot{n} &= -\lambda nm + \mu f - \delta n + \nu \\ \dot{m} &= -\lambda nm + \mu f + \delta f + \gamma f \\ \dot{f} &= \lambda nm - \mu f - \delta f - \gamma f\end{aligned} \tag{1.28}$$

where we denoted the number of enzyme-protein complexes with f. The rates λ and μ correspond to association and dissociation of the protein to the enzyme respectively. The rate γ is the rate of degradation of a protein bound to an enzyme. The rates ν and δ correspond to the birth and death process of the protein. The protein can decay as a free protein and while it is bound to the enzyme f

For the total number of enzymes M we obtain the conservation law $m + f = M$. We calculate the change in the total number of proteins that is given by the sum of free proteins and the number of proteins bound to the enzyme and obtain the differential equation

$$\frac{d}{dt}(n+f) = -\delta(n+f) + \nu - \gamma f \tag{1.29}$$

We simplify this equation by assuming that $\dot{f} = 0$. This corresponds to a constant occupation of the enzyme machinery. This occupation is given by

$$f = \frac{Kn}{1+Kn}M = \tilde{f}n \qquad \text{where} \qquad \tilde{f} = \frac{KM}{1+Kn}, K = \frac{\lambda}{\mu+\gamma+\delta} \tag{1.30}$$

In the time-interval Δt the number of free enzymes n and with that $\tilde{f}$ are constant. With the substitution

$$\tilde{\delta} = \frac{\delta(1+\tilde{f})+\gamma\tilde{f}}{1+\tilde{f}^2K^{-1}M^{-1}} \qquad \tilde{\nu} = \frac{\nu}{1+\tilde{f}^2K^{-1}M^{-1}} \tag{1.31}$$

we can rewrite (1.29) identically to the differential equation for birth and death (Methods Equation (1.2))

$$\dot{n} = -\tilde{\delta}n + \tilde{\nu} \tag{1.32}$$

with the well-known solution

$$n = n_0 \exp(-\tilde{\delta}t) + \frac{\tilde{\nu}}{\tilde{\delta}}(1 - \exp(-\tilde{\delta}t)) \tag{1.33}$$

For the production rate ν we take the degraded molecules δn to ensure particle conservation. In other words we modify the last rates in (Methods Equation 1.3) with (1.33) such that they depend on the number of degrading enzymes M in contrast to being constant.

1.3.3 Reversible Michealis-Menten kinetics

We consider Michaelis-Menten kinetics where an inactive substrate S_0 is activated to S_1 by an Enzyme E (Figure 2 c). We denote the complex of enzyme and inactive substrate with f and the number of free enzymes with m_E. We introduce a reverse rate γ_2 that describes the kinetics of spontaneous deactivation of S_1. The system of differential equations is

$$\begin{aligned}
\dot{S}_0 &= -\lambda m_E S_0 + (\mu + \gamma_1 + \delta_E) f - \delta_S S_0 - \gamma_1 f + \gamma_2 S_1 \\
\dot{m}_E &= -\lambda m_E S_0 + (\mu + \gamma_1 + \delta_S) f + \nu_E - \delta_E m_E \\
\dot{f} &= \lambda m_E S_0 - (\mu + \gamma_1 + \delta_S + \delta_E) f \\
\dot{S}_1 &= -\gamma_2 S_1 + \gamma_1 f - \delta_S S_1 + \nu_S
\end{aligned} \tag{1.34}$$

Here, λ and μ correspond to association and dissociation of the protein to the enzyme respectively. The rate of activation of S_0 being bound to E is denoted by γ_1. The birth and death processes of S_1 and E are denoted by ν_S, δ_S and ν_E, δ_E respectively. We assume again that the pre-steady-state phase of enzyme occupation is fast. In addition, the reversible channel prevents complete depletion of substrate. Thus $\dot{f} = 0$ is a good approximation. We obtain

$$f = \frac{KS_0}{1 + KS_0} E \qquad \text{where} \qquad K = \frac{\lambda}{\mu + \gamma_1 + \delta_E + \delta_S} \tag{1.35}$$

We next calculate the fixed point

$$f = \tilde{K}(E - f)(\tilde{S} - f) \qquad \tilde{K} = K \frac{\gamma_1 + \gamma_2 + \delta_S}{\gamma_2 + \delta_S} \qquad \tilde{S} = S \frac{\gamma_2}{\gamma_1 + \gamma_2 + \delta_S} \tag{1.36}$$

To approximate transient behavior toward this fixed point we linearize the ODE and obtain the effective rate

$$k = \gamma_1 \frac{f'(S_0^{eq})}{1 + f'(S_0^{eq})} + \gamma_2 + \delta_S \tag{1.37}$$

for its kinetics. In the above equation denoted the derivative of $f(S_0)$ with respect to S_0 at the fixed point S_0^{eq} as $f'(S_0^{eq})$. With this approximation for transient rate towards equilibrium we obtain the time dependent amount of S_0 with the initial condition S_0^{ivp}

$$S_0(t) = S_0^{ivp} \exp(-kt) + S_0^{eq}(1 - \exp(-kt)) \tag{1.38}$$

1.3.4 Enzyme activities controlled by external factors

We consider an enzyme that is activated by receptors and deactivated by a repressive cofactor (Figure 2d). The unoccupied Receptor R catalyses the formation of the activated enzyme E_1 from the inactivated enzyme E_0. This includes formation of the complex f that consists of E_0 and R and in a successive step the formation of E_1 and the release of R. The reverse reaction is governed by the enzyme D where the complex of D and E_1 is denoted by g. To model these two processes we use the classical Michaelis-Menten-Modell for both enzymes. For R we denote λ_1 as the rate of formation of the enzyme substrate complex f, μ_1 as the rate governing the stability of the complex and γ_1 as the catalysis rate. The ODEs for the reaction catalyzed by R are

$$\begin{aligned}
\dot{m}_R &= -\lambda_1 E_0 m_R + (\delta_E + \mu_1 + \gamma_1) f - \delta_R m_R + \nu_R \\
\dot{E}_0 &= -\lambda_1 E_0 m_R + (\delta_R + \mu_1 + \gamma_1) f - \delta_E E_0 - \gamma_1 f + \gamma_2 g \\
\dot{f} &= -(\mu_1 + \gamma_1 + \delta_E + \delta_R) f + \lambda_1 E_0 m_R
\end{aligned} \tag{1.39}$$

Where we denoted the birth and death processes of E and R with ν_E, δ_E and ν_R, δ_R respectively. For D we denote λ_2 as the rate of formation of the enzyme substrate complex g , μ_2 as the rate governing the stability of the complex and γ_2 as the catalysis rate. The ODEs for the reaction catalyzed by R are

$$\begin{aligned}
\dot{m}_D &= -\lambda_2 E_1 m_D + (\delta_E + \mu_2 + \gamma_2) g - \delta_D m_D + \nu_D \\
\dot{E}_1 &= -\lambda_2 E_1 m_D + (\delta_D + \mu_2 + \gamma_2) g - \delta_E E_1 + \gamma_1 f - \gamma_2 g + \nu_E \\
\dot{g} &= -(\mu_2 + \gamma_2 + \delta_E + \delta_D) g + \lambda_2 E_1 m_D
\end{aligned} \tag{1.40}$$

From these equations we obtain mass conservation laws for the passive and active enzymes

$$\begin{aligned}
\frac{d}{dt}(E_0 + f) &= -\delta_E (E_0 + f) - \gamma_1 f + \gamma_2 g \\
\frac{d}{dt}(E_1 + g) &= -\delta_E (E_1 + g) + \gamma_1 f - \gamma_2 g + \nu_E
\end{aligned} \tag{1.41}$$

We assume that $\dot{f} = \dot{g} = 0$ and obtain the Michaelis Menten type equations for the overall number of Receptors $R = m_R + f$ and the overall number of inhibitors $D = m_D + g$

$$\begin{aligned}
f &= \frac{K_1 E_0}{1 + K_1 E_0} R \qquad K_1 = \frac{\lambda_1}{\mu_1 + \gamma_1 + \delta_E + \delta_R} \\
g &= \frac{K_2 E_1}{1 + K_2 E_1} D \qquad K_2 = \frac{\lambda_2}{\mu_2 + \gamma_2 + \delta_E + \delta_D}
\end{aligned} \tag{1.42}$$

The fixed point is obtained by setting (1.41) to zero and solving the system of (1.41)(1.42) applying the particle conservation law $E = E_0 + E_1 + f + g$. The iteration for the fixed point is

$$\begin{aligned}
E_0 &= E\left(1 + \frac{f + g + E_1}{E_0}\right)^{-1} \\
f &= \frac{K_1 E_0}{1 + K_1 E_0} R \\
g &= \frac{1}{\gamma_2}(\delta_E (E_0 + f) + \gamma_1 f) \\
E_1 &= \frac{g}{K_2 D}(1 + K_2 E)
\end{aligned} \tag{1.43}$$

To approximate the transient behavior toward the fixed point we use the linearization

$$k = \delta_E + \gamma_1 \frac{f'(E_0)}{1 + f'(E_0)} + \gamma_2 \frac{g'(E_1)}{1 + g'(E_1)} \tag{1.44}$$

where we denote the derivatives of $f(E_0)$ and $g(E_1)$ as $f'(E_0), g'(E_1)$. The solution of the linearized differential equation in proximity of the fixed point is

$$E_0(t) = E_0^{ivp} \exp(-kt) + E_0^{eq}(1 - \exp(-kt)) \tag{1.45}$$

where E_0^{ivp} is the initial condition and E_0^{eq} is the number at the fixed point.

1.3.5 Enzyme Inhibitor

We describe the activation of substrate S_0 by an enzyme R forming a complex f with association and dissociation rates λ_1 and μ_1. The complex produces the active substrate S_1 with rate γ_1. The active substrate can form a homo-dimer S_2 with association and dissociation rates λ_3 and μ_3. We include a species X in the reaction that can inhibit the enzyme by forming a complex g with the enzyme (Figure 2 e). The occupation of the enzyme by the inhibitor is determined by the association and dissociation rates λ_2, μ_2. We denote the number of unbound inhibitors as m_X.

$$\begin{aligned}
\dot{m}_x &= -\lambda_2 f m_x - \delta_x m_x + \nu_x + (\mu_2 + \delta_R + \delta_S) g \\
\dot{g} &= \lambda_2 f m_x - (\mu_2 + \delta_R + \delta_S + \delta_x) g \\
\dot{f} &= \lambda_1 S_0 m_R - \lambda_2 f m_x - (\mu_1 + \gamma_1 + \delta_R + \delta_S) f + (\mu_2 + \delta_x) g \\
\dot{m}_R &= -\delta_R m_R + \nu_R - \lambda_1 S_0 m_R + (\mu_1 + \gamma_1 + \delta_S) f + \delta_S g \\
\dot{S}_0 &= -\delta_S S_0 - \lambda_1 S_0 m_R + (\mu_1 + \delta_R) f + \gamma_2 S_1 + \delta_R g \\
\dot{S}_1 &= -(\gamma_2 + \delta_S) S_1 + \nu_S + \gamma_1 f - 2\lambda_3 S_1^2 + 2(\mu_3 + \delta_S) S_2 \\
\dot{S}_2 &= \lambda_3 S_1^2 - (\mu_3 + 2\delta_S) S_2
\end{aligned} \tag{1.46}$$

The birth and death processes of S_1 and E are denoted by ν_S, δ_S, ν_R, δ_R and ν_x, δ_x respectively. We obtain the differential equation for the total number of inactive and active substrate

$$\begin{aligned}
\frac{d}{dt}(f + g + S_0) &= -\delta_S (f + g + S_0) - \gamma_1 f + \gamma_2 S_1 \\
\frac{d}{dt}(S_1 + 2S_2) &= -\delta_S (S_1 + 2S_2) + \nu_S + \gamma_1 f - \gamma_2 S_1
\end{aligned} \tag{1.47}$$

From (1.46) we can find the particle conservation laws for the total number of inhibitor, the total number of enzyme and the total number of substrate

$$\begin{aligned}
X &= m_x + g \\
R &= m_R + f + g \\
S &= S_0 + S_1 + 2S_2 + f + g
\end{aligned} \tag{1.48}$$

We next introduce new variables to rewrite the system of equations into the previously discussed heterodimer-case (Section 1.3.1)

$$\begin{gathered}
K_1 = \frac{\lambda_1}{\mu_1 + \gamma_1 + \delta_R + \delta_S} \quad K_2 = \frac{\lambda_2}{\mu_2 + \delta_R + \delta_S + \delta_X} \quad K_3 = \frac{\lambda_3}{\mu_3 + 2\delta_S} \quad \Delta K = \frac{\delta_R + \delta_S}{\mu_1 + \gamma_1 + \delta_R + \delta_S} \\
\tilde{X} = X(1 - \Delta K) - (1 - q)R \\
q = \frac{1}{1 + \Delta K K_2 X} \\
\tilde{R} = qR \qquad \tilde{K}_2 = qK_2
\end{gathered} \tag{1.49}$$

As argued in (Section 1.3.3), we assume that the enzymes are in a steady state corresponding to the relation $\dot{f} = \dot{g} = 0$. We then find

$$g = \frac{K_2 f}{1 + K_2 f} X \tag{1.50}$$

$$f + \Delta K g = K_1 S_0 m_R \tag{1.51}$$

We introduce the new variable

$$d = \tilde{R} - \frac{f}{r} \quad \text{where} \quad r = \frac{K_1 S_0}{1 + K_1 S_0} \tag{1.52}$$

to obtain an equation that can be solved using (1.22)

$$d = \tilde{K}_2 r (\tilde{R} - d)(\tilde{X} - d) \tag{1.53}$$

We obtain the fixed point by iterating the equations

$$\begin{aligned}
d &= \frac{1}{\tilde{K}_2 r} \bar{d}(\tilde{K}_2 r \tilde{R}, \tilde{K}_2 r \tilde{X}) \\
f &= r(\tilde{R} - d) \\
S_1 &= \frac{\gamma_1}{\gamma_2} f + \frac{\delta_S}{\gamma_2}(f + g + S_0) \\
S_2 &= K_3 S_1^2 \\
S_0 &= S\left[\left(1 + \frac{\delta_S}{\gamma_2}\right)\left(1 + \frac{f+g}{S_0}\right) + \frac{\gamma_1}{\gamma_2}\frac{f}{S_0} + 2K_3 \frac{S_1^2}{S_0}\right]^{-1}
\end{aligned} \tag{1.54}$$

We linearize the differential equation close to the fixed point and obtain the rate constant k for the transient kinetics toward the fixed point.

$$\begin{aligned}
k &= \delta_S + \gamma_1 \frac{f'(S_0)}{1 + f'(S_0) + g'(S_0)} + \gamma_2 \frac{1}{1 + 4K_3 S_1} \\
f'(S_0) &= \frac{r'(S_0)}{r}\left(f - \frac{1}{\tilde{K}_2}\frac{d^2}{\tilde{R}\tilde{X} - d^2}\right) \\
g'(S_0) &= \frac{K_2 X}{\left(1 + K_2 f\right)^2} f'(S_0) \\
r'(S_0) &= K_1 (1 - r)^2
\end{aligned} \tag{1.55}$$

where we denote derivatives as primed functions. With this approximation for transient rate towards equilibrium and the values at the fixed point we obtain the time dependent amount of S_0 with the initial condition S_0^{ivp}

$$S_0(t) = S_0^{ivp} \exp(-kt) + S_0^{eq}(1 - \exp(-kt)) \tag{1.56}$$

1.3.6 Noise in chemical reactions

All kinetic rate constants in the reactions discussed above represent stochastic Poisson processes. However, we merely offered deterministic solutions for the steady state and the transient processes towards it. If the number of molecules per species in the reaction is high, these deterministic solutions are sufficient for the description of the reaction in the context of the GRN. For networks that may be

affected by noise of reactions, CaiNet offers a strongly simplified representation of the exact noisy fluctuations in steady state levels. The steady state level is assumed to be the expectation value of a Poisson distribution. CaiNet draws a random number for this distribution, that that represents the new steady state level transient processes are converging to. In the following we discuss the validity of this procedure for the homo-dimer case.

To calculate the noise due to stochastic fluctuations in the molecule level of dimers we write down the Kolmogorov equations for the probability to find n Dimers at time t

$$\dot{p}(n) = -\lambda(N-2n)(N-1-2n)\,p(n) + \mu(n+1)\,p(n+1) \tag{1.57}$$

where N is the total number of monomers. In equilibrium this equation relates p(n) and p(n+1) by

$$\mu(n+1)\,p(n+1) = \lambda(N-2n)(N-1-2n)\,p(n) \tag{1.58}$$

With this recursion formula we can write down the absolute value for the probability

$$p_n = \frac{z_n K^n}{Z} \quad \text{where} \quad z_n = \frac{N!}{(N-2n)!n!}, Z = \sum_n z_n K^n \tag{1.59}$$

We sum over (1.58) and obtain an equation for the expectation value of n, $\bar{n}$

$$\bar{n} = K(N-2\bar{n})(N-1-2\bar{n}) + 4K\cdot(\overline{n\cdot n} - \bar{n}^2) \approx K(N-2\bar{n})(N-2\bar{n}) \tag{1.60}$$

where we assumed that $4K\cdot(\overline{n\cdot n} - \bar{n}^2) << \bar{n}$. We calculate the variance of n with

$$Var(n) = K\frac{d\bar{n}}{dK} = \frac{\bar{n}}{1+4\sqrt{K\bar{n}}} \tag{1.61}$$

Supplementary Table 1: Parameters for the elements

Element Name	Symbol	Meaning
Dimerization	λ	Association rate of the two monomers
	μ	Inverse Half Life of the dimer
Targeted Degradation	λ	Association rate of the degradation complex to the protein of interest
	μ	Dissociation rate of the protein from the degradation complex without degradation happening
	γ	Rate of degradation of the protein while being bound to the degradation complex
Activation and Inhibition by enzymes	λ_1	Association rate of the activating enzyme to the protein of interest
	μ_1	Dissociation rate of the protein from the activating enzyme without activation happening
	γ_1	Activation rate while bound to the activating enzyme
	λ_2	Association rate of the deactivation enzyme to the protein of interest
	μ_2	Dissociation rate of the protein from the deactivating enzyme without deactivation happening
	ν_2	Deactivation rate while bound to the deactivating enzyme
Enzyme inhibitor	λ_1	Association rate of the activating enzyme to the protein of interest
	μ_1	Dissociation rate of the protein from the activating enzyme without activation happening
	γ_1	Activation rate while bound to the activating enzyme
	λ_2	Association rate of the inhibitor to the enzyme substrate complex
	μ_2	Dissociation rate of the inhibitor from the enzyme substate complex
	γ_2	Constant deactivation rate of the protein of interest
	λ_3	Association rate of the two monomers
	μ_3	Inverse Half life of the dimer
Michaelis menten	λ_1	Association rate of the enzyme to the protein of interest
	μ_1	Dissociation rate of the protein from the enzyme without activation happening
	γ_1	Systhesis rate while bound to the activating enzyme
	γ_2	Rate of the inverse process

2 Simulated Examples

2.1 Validation: Pulse activated auto feedback loop

To validate our hybrid stochastic deterministic algorithm, we compared CaiNet with the Gillespie algorithm. We chose an example where minimal stochastic fluctuations in protein synthesis, as well as in switching between on and off states in the two state model, can emerge in a different outcome with respect to de-/activation of the auto feedback loop.

We considered a single gene. This gene is activated by an external factor that is only present for a defined pulse-length. The gene produces transcription factors, which dimerize and activate their own transcription. Kinetic parameters have been chosen such that this autocatalysis can collapse resulting in a expression level close to zero. We investigate the probability to observe this collapse depending on the pulse length of the activating external factor.

2.1.1 Gillespie simulation with a custom programed Excel-Matlab-UI

To validate our algorithm, we used the Gillespie algorithm. Currently many tools implement the Gillespie Method (the direct method as well as the sped up method) and include user-interfaces for entering the coupled reaction system [1-4]. However, for the special task of simulating gene networks we wrote a custom implementation in Matlab that is complemented by a rudimentary user-interface in excel. (See Code Availability Statement)

The Gillespie method invokes two steps: 1) Calculating the point in time of the next reaction event from all reaction rates 2) Deciding which reaction event actually takes place. These two steps can become computationally expensive since they invoke iterations through all reactions of the system. Therefore, algorithms have been developed that group reactions and use two random numbers to first select a subgroup and then the reaction inside the subgroup [5]. In doing so, they avoid iterating trough every single reaction.

For the case of gene regulatory networks, we perform a heuristic grouping based on reaction times scales to form three subgroups: 1) birth and death processes 2) Chemical reactions 3) Binding of TFs to promoters. This choice of groups reduces computation time during reaction type selection.

In the pulse-activated auto feedback loop described in (Supplementary Information Section 2.1 and Figure 1 Panel c) The reactions for binding/unbinding of TFs to the promoter are

$$\begin{aligned} &\text{Promoter} + \text{TF} \rightleftharpoons \text{Promoter-TF} \\ &\text{Promoter} + \text{ExtTF} \rightleftharpoons \text{Promoter-ExtTF} \end{aligned} \tag{1.62}$$

Where the activation of the promoter leads to production of TF. The Birth and death reactions are

$$\begin{aligned} &\text{Promoter-TF} \longrightarrow \text{Promoter-TF} + \text{TF} \\ &\text{Promoter-TF} \longrightarrow \text{Promoter-TF} + \text{TF} \\ &\text{TF} \longrightarrow \varnothing \end{aligned} \tag{1.63}$$

The external TF extTF is not subject to any dynamics, since its levels are given as boundary conditions. The chemical reaction rates are

$$\begin{aligned} &2TF \rightleftharpoons TF_2 \\ &TF_2 \longrightarrow TF + \varnothing \end{aligned} \tag{1.64}$$

With our Excel-UI this set of reactions can be formulated in a simplified way, since chemical reactions of birth and death-processes as well as state switching of promoters are automatically compiled into a set of reactions that can be processed by the Gillespie-formalism.

The reaction parameters are:

Supplementary Table 2: Parameters for the autofeedback-loop

TFEXT-Promoter Association	$\lambda_{Input} = 1e-3$
TFEXT-Promoter-Dissociation	$\mu_{input} = 5e-1$
TF-Promoter Association	$\lambda_{TF} = 2e-3$
TF-Promoter-Dissociation	$\mu_{TF} = 7e-2$
TF Production	$\nu = 25e-2$
TF Degradation	$\delta = 1e-3$
TF Dimerization-Rate	$\kappa_1 = 1e-2$
TF Dimerization half-life	$\kappa_2 = 1$

2.1.2 Ordinary differential equations

As a next step, we compared the Gillespie algorithm with the solution of the ordinary differential equations of the pulse-activated auto feedback loop. The ODEs are (parameters are defined in Supplementary Table 2)

$$\begin{aligned}\dot{n} &= \frac{K_{Ext}\cdot a + K_{TF}}{1 + K_{Ext}\cdot a + K_{TF}}\nu - \delta n - 2\kappa_1 n^2 + 2\kappa_2 m + 2\delta m \\ \dot{m} &= \kappa_1 n^2 - \kappa_2 m - 2\delta m\end{aligned} \tag{1.65}$$

where

$$a = \begin{cases} 1, 0 \le t \le t_0 \\ 0, t_0 \le t \end{cases} \tag{1.66}$$

We first numerically calculated the solution from t to t_0 with $a = 1$ and then used the values $n(t_0)$ as an initial value problem for the ODE with $a = 0$. If the number of monomers n decayed to zero we interpreted this as staying in the off-state. We chose simulation parameters identical to the Gillespie parameters (Supplementary Table 2)

$$K_{Ext} = 2e-3; K_{TF} = \frac{2}{70}; \nu = 0.25; \delta = 1e-3; \lambda = 1e-2; \mu = 1 \tag{1.67}$$

2.1.3 CaiNet

Finally, we compared the Gillespie algorithm with the stochastic-deterministic algorithm of CaiNet for the pulse-activated auto feedback loop.

We implemented the external TF as an input element. The TF-gene element was chosen to be an OR-Logic element since it obtains activating inputs from the external TF and from its own TF-Dimers. The TF-Dimers were implemented by the dimerization element which was connected to the output of the TF-gene element.

By default, stochasticity in this setup is only contributed by the OR-Logic element which stochastically switches between on/off states. In addition, stochasticity of the birth-process as well as noise in dimer

levels can be switched on with the formalism described in (Methods Section 1.1.2 and Supplementary Information Section 1.3.6). In the Gillespie algorithm Dimers immediately take action in activating the promoter one they have been formed. In contrast, the dimer-population in CaiNet is updated with a delay corresponding to the time-step Δt .

For CaiNet we chose the same parameters as for the Gillespie simulation. To test whether the step size of CaiNet has an impact on the simulation outcome, we ran simulations for the values 0.1s,1s,10s,100s (Supplementary Supplementary Figure 4). We found that the step sizes 0.1s, 1s and 10s resemble each other reasonably well, while the step size of 100s visibly underestimated the switching probability for short pulse durations. Comparing this step size to the shortest applied pulse duration 300s and to the order of magnitude of the gene switching rates 1e-2 this result is to be expected.

2.2 Circadian clock

2.2.1 Comparing mechanisms of repression of Bmal/Clock targets by Per/cry

In our simplified picture of Bmal/Clock target gene expression, the gene is transcribed as long as Bmal/Clock occupy its promoter. Per/Cry acts as a repressor of Bmal/Clock target genes by either directly inhibiting the Bmal/Clock heterodimer through formation of a Bmal/Clock-Per/Cry tetramer, or by blocking promoters of Bmal/Clock target genes such that Bmal/Clock cannot bind. Currently there is no experimental evidence to support one of these two variants of Bmal/Clock repression. Here we compare the two scenarios to discuss whether they affect modelling of the circadian clock. We start with tetramer formation and compare our results to the two state model with repression calculated in (Supplemtary Information 1.2.5)

We assume that Bmal/Clock A and Per/Cry R heterodimers are constant. The number of tetramers d in equilibrium is given by

$$d = K(A-d)(R-d) \tag{1.68}$$

Where K is the equilibrium constant of tetramer formation. To better point our reasoning we rewrite this equation and obtain

$$d = \frac{KAR}{1+KA+KR-Kd} \tag{1.69}$$

We assume that $Kd \ll 1+KA+KR$. This means that the number of heterodimers is either limited by one of the monomers or the equilibrium constant of dimerization is small. With this assumption we obtain

$$d = \frac{KAR}{1+KA+KR} \tag{1.70}$$

The number of tetramers d equals the amount of inhibited Bmal/Cry heterodimers. We obtain the amount of active Bmal/Cry heterodimers A_1 with the law of particle conservation

$$A_1 = A-d = A\frac{1+KA}{1+KR+KA} \tag{1.71}$$

The probability of finding a target promoter occupied by Bmal/Clock is

$$p = \frac{K_1A_1}{1+K_1A_1} = \frac{K_1A(1+KA)}{(1+K_1A)(1+KA)+KR} \tag{1.72}$$

where K_1 is the product of the arrival rate and the binding time of Bmal/Clock. Comparing this to the result from (1.14)

$$p = \frac{K_1A}{1+K_1A+K_2R} \tag{1.73}$$

we find that the two equations are the same if $KA \ll 1$.

From this we conclude that both scenarios yield qualitatively same behavior. Thus, we decided to restrict ourselves to one scenario, namely exclusive tetramer formation at the promoter.

2.2.2 Deterministic model of the internal loop of the circadian clock

We here investigate the internal loops of the circadian clock consisting of a negative lobe mediated by Rev and a positive lobe mediated by Ror (Supplementary Figure 5 panel a). Both the Rev and Ror genes are activated by the Bmal/Clock heterodimer. The Rev and Ror TFs in turn independently repress and activate the transcription of clock and bmal genes. We added two constant signals N_1 and N_2 to the system. These activate the transcription of Bmal and Clock. Using the function for dimer formation $\bar{d}(u,v)$ defined in (1.22), the system of ODEs of the circadian clock is given by

$$\begin{aligned} \dot{ROR} &= -\delta_1 ROR + \nu_1 A \\ \dot{REV} &= -\delta_2 REV + \nu_2 A \\ \dot{BMAL} &= -\delta_3 BMAL + \nu_3 B_1 \\ \dot{CLOCK} &= -\delta_3 CLOCK + \nu_4 B_2 \\ Bmal\,/\,Clock &= K_4^{-1} \cdot \bar{d}(K_4 Bmal, K_4 Clock) \end{aligned} \tag{1.74}$$

where A is the probability of a Bmal/Clock target gene to be active, B_1 is the probability of the Bmal Promoter to be active and B_2 is the probability of the Clock promoter to be active.

$$\begin{aligned} A &= \frac{K_1 \cdot BMALCLOCK}{1 + K_1 \cdot BMALCLOCK} \\ B_1 &= \frac{K_3 \cdot ROR + K_3 \cdot N_1 + K_3^2 \cdot ROR \cdot N_1}{(1 + K_3 \cdot ROR) \cdot (1 + K_3 \cdot N_1)} \cdot \frac{1}{1 + K_2 \cdot REV} \\ B_2 &= \frac{K_3 \cdot ROR + K_3 \cdot N_2 + K_3^2 \cdot ROR \cdot N_2}{(1 + K_3 \cdot ROR) \cdot (1 + K_3 \cdot N_2)} \cdot \frac{1}{1 + K_2 \cdot REV} \end{aligned} \tag{1.75}$$

We now renormalize this system of equations with the variables

$$\begin{aligned} &ROR = \frac{\nu_1}{\delta_1} x \quad REV = \frac{\nu_2}{\delta_2} y \quad BMAL = \frac{\nu_3}{\delta_3} z_1 \quad CLOCK = \frac{\nu_4}{\delta_3} z_2 \\ &q_1 = \frac{\nu_1}{\delta_1 N_1} \quad q_2 = \frac{\nu_1}{\delta_1 N_2} \quad q_3 = \frac{K_1}{K_4} \\ &n_1 = \frac{\nu_1}{\delta_1} K_3 \quad n = \frac{\nu_2}{\delta_2} K_2 \quad n_3 = \frac{\nu_3}{\delta_3} K_4 \quad n_4 = \frac{\nu_4}{\delta_3} K_4 \end{aligned} \tag{1.76}$$

and obtain

$$\begin{aligned} \dot{x} &= -\delta_1 (x - A) \\ \dot{y} &= -\delta_2 (y - A) \\ \dot{z}_1 &= -\delta_3 (z_1 - B_1) \\ \dot{z}_2 &= -\delta_3 (z_2 - B_2) \\ A &= 1 - \frac{1}{1 + q_3 \bar{d}(n_3 z_1, n_4 z_2)} \\ B_{1,2} &= \frac{1}{1 + n_2 y} \left(1 - \frac{1}{1 + K_3 N_{1,2}} \frac{1}{1 + n_1 x} \right) \end{aligned} \tag{1.77}$$

2.2.3 Parameterizing the circadian clock

When setting up the network of the circadian clock, we observed that it does not oscillate for arbitrary sets of parameters. We here show a formalism that provides rules on how to choose parameters for oscillations of the circadian clock network. We started with the inner clock consisting of Bmal Clock Rev Ror. Our starting point was the ODEs from (1.74). To simplify the calculation, we assumed that occupation of promoters by these TFs is small. This means that $K_{TF} \cdot n_{TF} << 1$ for BMAL/CLOCK and ROR. In addition, we assumed that the number of the BMAL/CLOCK Dimer d can be calculated by the number of the two monomers M_1 and M_2 by $d = K_4 M_1 M_2$ with the equilibrium constant K_4. With these assumptions we obtain the simplified probabilities of active promoters from (1.74)

$$\begin{aligned} A &= K_1 K_4 BMAL \cdot CLOCK \\ B_1 &= \frac{K_3 N_1 (1 + ROR / N_1)}{1 + K_2 REV} \\ B_2 &= \frac{K_3 N_2 (1 + ROR / N_2)}{1 + K_2 REV} \end{aligned} \tag{1.78}$$

An empirical observation when applying guessed parameters in our simulations showed that

$$\delta_1 >> \delta_2 \tag{1.79}$$

supports oscillations. We further assume that the timescales of δ_1 and δ_3 are well separated such that the adiabatic approximation

$$\begin{aligned} BMAL &= \nu_3 / \delta_3 B_1 \\ CLOCK &= \nu_4 / \delta_3 B_2 \end{aligned} \tag{1.80}$$

holds. We then use normalized variables from (1.76) to write down the normalized and simplified system of ODEs

$$\begin{aligned} \dot{x} &= -\delta_1 (x - A) \\ \dot{y} &= -\delta_2 (y - A) \\ A &= s \frac{(1 + q_1 x)(1 + q_2 x)}{(1 + n y)^2} \\ s &= K_1 K_4 K_3^2 N_1 N_2 \frac{\nu_3 \nu_4}{\delta_3^2} \end{aligned} \tag{1.81}$$

For clarity we here only look at the case $q_2 = n$. Our aim is now to create a fixed point with repelling dynamics such that the circuit never adopts a constant stable value. In our system we can have a maximum of two fixed points. Since one of the two possible fixed points is negative we are left with one fixed point x_0 that is calculated by

$$x_0 (1 + n x_0) = s (1 + q_1 x_0) \tag{1.82}$$

The solution of this quadratic equation for x_0 is given by

$$x_0 = \frac{q_1 s - 1 + \sqrt{(q_1 s - 1)^2 + 4sn}}{2n} \tag{1.83}$$

Knowing the fixed point we can calculate the condition for repelling eigenvalues i.e. eigenvalues that have a positive real part. A sufficient condition for at least one positive real part is that the trance of the matrix of the linearized system, is larger than zero. The diagonal elements are

$$\begin{aligned} m_{11} &= \delta_1 \frac{q_1 s - 1}{1 + n x_0} \\ m_{22} &= -\delta_2 \frac{1 + 3 n x_0}{1 + n x_0} \end{aligned} \tag{1.84}$$

We obtain the necessary condition for positive real part of the eigenvalues

$$\frac{\delta_2}{\delta_1} < \frac{q_1 s - 1}{1 + 3 n x_0} \tag{1.85}$$

In particular, this means

$$q_1 s - 1 > 0 \tag{1.86}$$

2.2.4 Estimation of the frequency

We have constructed a repelling fixed point that ensures that the circadian network exhibits periodical dynamics. To tune the frequency of the clock to approximately 24h we rewrite the system of ODEs to obtain a classical lienard system. We apply the transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{(\delta_1 - \delta_2)} \begin{pmatrix} -\delta_2 & \delta_1 \\ \delta_2 - \delta_1 & \delta_1 - \delta_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.87}$$

to obtain the system

$$\begin{aligned} \dot{u} &= -\frac{\delta_1 \delta_2}{\delta_1 - \delta_2} v \\ \dot{v} &= (\delta_1 - \delta_2)(u - A) - (\delta_1 + \delta_2) v \end{aligned} \tag{1.88}$$

We eliminate v to obtain a nonlinear second order ODE with nonconstant coefficients

$$\ddot{u} = -\delta_1 \delta_2 (u - A) - (\delta_1 + \delta_2) \dot{u} \tag{1.89}$$

Linearizing A yields with respect to $\dot{u}$ yields a Lienard type ODE

$$\ddot{u} = -f(u) - g(u) \dot{u} \tag{1.90}$$

with $f(u) = \delta_1 \delta_2 (u - A(u,u))$. The frequency of oscillation can be approximated by

$$\omega^2 = f'(x_0) \tag{1.91}$$

To obtain a frequency of one Day, the relation

$$\delta_1 \delta_2 = \omega^2 \frac{1 + n x_0}{2 n x_0 - q_1 s + 1} \tag{1.92}$$

has to hold. Plugging this into (1.85) yields

$$\delta_1 > \omega \sqrt{\frac{1+nx_0}{2nx_0 - q_1 s + 1} \cdot \frac{1+3nx_0}{q_1 s - 1}} \tag{1.93}$$

2.2.5 The internal loop of the circadian clock in an autonomous oscillator

We used the parameters determined above to create oscillations in the internal loop without per and cry. We found that NFYA can modulate the phase of this internal loop.

The core clock network was set up according to the entry in the KEGG database [6] omitting the transcription factor Dec. We assumed that Bmal and Clock expression is regulated by the same promoter [7] and that this promoter has a constant ground activation Nfy [8].

We set up the network in CaiNet using OR-logic elements for ROR, REV, BMAL and CLOCK. We used the dimer element to include the dimerization of Bmal and Clock. The elements have been connected amongst each other according to (Supplementary Table 3)

The internal loop shows a periodical oscillation with a period of ~24h. We varied the level of NFY using a pulsed input signal. This led to a Phase shift in oscillation of circadian clock genes.

We tested whether per can induce a Phase shift, while the frequency of the oscillation of the circadian clock is preserved. For this, we added Per/Cry as the Stat3/enzyme inhibition element. (ref. calculation) We modulated the inhibition of Per/Cry dimers with an Input element. We observed that a constant shift in Per pulsing resulted in a shift of the maxima of the oscillations. Fig. 5d.

Supplementary Table 3: Parameters and connections in the circadian Network

Species	Interaction	Rate	Birth/Death
ROR	BMAL/CLOCK	$\lambda = 5e-4 \quad \mu = 1$	$\nu = 33.3 \quad \beta = 1 \quad \delta = 3.67e-4$
	Per/Cry	$\lambda = 1e-4 \quad \mu = 1e-2$	
REV	BMAL/CLOCK	$\lambda = 5e-4 \quad \mu = 1$	$\nu = 1.96 \quad \delta = 3.67e-4$
	Per/Cry	$\lambda = 1e-4 \quad \mu = 1e-2$	
BMAL	ROR	$\lambda = 1e-4 \quad \mu = 1$	$\nu = 10 \quad \beta = 1 \quad \delta = 0.001$
	REV	$\lambda = 1e-2 \quad \mu = 1$	
	NFY	$\lambda = 1e-4 \quad \mu = 1$	
CLOCK	ROR	$\lambda = 1e-4 \quad \mu = 1$	$\nu = 10 \quad \beta = 1 \quad \delta = 0.001$
	REV	$\lambda = 1e-2 \quad \mu = 1$	
	NFY	$\lambda = 1e-4 \quad \mu = 1$	
BMAL/CLOCK	Bmal Clock	$\lambda = 1e-2 \quad \mu = 1$	N/A
PER Cry	BMAL/CLOCK	$\lambda = 5e-4 \quad \mu = 1$	$\nu = 2 \quad \delta = 5e-4$
	Per/Cry Dimer	$\lambda = 1e-4 \quad \mu = 1e-2$	
Per/Cry	Per/Cry	$\lambda_1 = 1e-3 \quad \mu_1 = 1 \quad \gamma_1 = 1$ $\lambda_2 = 1e-3 \quad \mu_2 = 1 \quad \gamma_2 = 1$ $\lambda_3 = 0.1 \quad \mu_3 = 0.001$	N/A

2.3 Pluripotency Network

2.3.1 Set up of the pluripotency network deduced from experiments

We built up the pluripotency network by searching the TRRUST database [9] and the KEGG database [6] for functional interactions between TFs. In accordance with [10] we separated the network in layers that communicate with each other. The first layer consists of the extracellular signals Lif, TGFβ, Bmp, Fgf4 and Wnt. Extracellular signals are transduced by a signal transduction layer that controls the activity of corresponding signaling transcription factors and includes crosstalk between extracellular signals. Transcription factors from the signal transduction layer regulate the core pluripotency gene regulatory network consisting of Nanog, Sox2 and Oct4, as well as markers of differentiation including Runx2, Pax6 and Gata6 in the last layer. Between the core network and the layer of differentiation markers is a feedback layer, whose transcription factors are controlled by positive and negative feedback loops of the two neighboring layers.

We first describe the layer of extracellular signals and their immediate targets. The cytokine Lif activates the Lif-Receptor, which in turn activates Janus kinases (Jak) that phosphorylate Stat3 monomers [11]. Stat3 dimerizes and activates expression of the downstream targets Socs and Klf4 [11, 12]. Socs inhibits the activation of Stat3 and thereby downregulates the response of the stat pathway to an external signal by Lif [11, 13]. Klf4 is associated with naive pluripotency and activates transcription of the core pluripotency factor Nanog [14]. The Lif receptor in addition activates the Ras/Raf cascade resulting in Erk activation [15].

Both the TGFβ-pathway and the BMP-pathway activate Smad TFs. TGFβ activates Smad2/3 and thereby the core pluripotency factor Nanog [16]. BMP activates Smad 1/5/8, which activates ID, Dusp and Runx2 [17]. The ID TF inhibits differentiation by inhibition of TCF3 [18], which represses the core pluripotency network [10]. Dusp inhibits the activity of Erk1/2. The Runx2 TF is a marker for mesoderm activation.

The fgf protein binds Fgf receptors, which activate the Ras/Raf cascade and thereby the ERK kinease [19]. ERK is responsible inhibition of the TFs Klf4 and TCF3 [20, 21]. In addition, it supports the auto feedback-loop of the endoderm differentiation marker Gata6 [22].

The WNT signaling pathway inhibits degradation of ctnnb1 by the GSK kinease. In turn ctnnb1 represses transcription of Tcf3 [23]. Tcf3 is also repressed by ID [18]. The TF Tcf3 shuts down the core pluripotency network by repressing transcription of the Oct4 activated mir302a [23] and Nanog [24]. The micro RNA mir302a inhibits repression of Oct4 by Nr2f2 [23, 25, 26].

Second, we set up the core pluripotency network. In our core network based on [27]., Nanog activates Sox2, which in turn activates Oct4 and represses itself [28]. For Oct4 we employed a complex promoter including cooperation of the TFs Sox2 and Oct4 [29]. Oct4 further activates Nanog.

Third, we describe the pathways into differentiation of the three germ lines. We identified differentiation markers that we could link to the core pluripotency network. These were Runx2 for mesoderm [30], Pax6 for ectoderm [31] and Gata6 for endoderm [22]. We found that none of these markers are direct targets of the core pluripotency fators, but all are activated by an interposed feedback layer.

Nanog activates the Snai1 transcription factor [32], which represses both its own transcription and transcription of the mesoderm differentiation marker Runx2 [33]. Sox2 activates the TF Six3 [34]. Six3 and the ectoderm differentiation marker Pax6 mutually activate each other [35, 36]. Oct4 forms a positive feedback loop with Sall4 [37], which represses the endoderm differentiation marker Gata6 [38]. Oct4 further activates transcription of Pax2 [39]. Pax2 and the ectoderm differentiation marker Pax6 mutually repress each other [40].

2.3.2 Set up of the pluripotency network in Cainet

We implemented all components of the Extracellular input layer as input elements with rectangular pulsed signals. The Stat pathway was implemented as the Stat-buildung element where the complex out of Jak-Kinease and Lif Receptor was identified as R and the Socs TF was assigned to the inhibitor. The Smads where implemented using the reversible michaelis menten element, ERK was implemented using the Forward/backward michaelis menten element where the forward reaction is controlled by fgf and the reverse reaction is controlled by Dusp. In the Wnt pathway is the element for GSK3 servs as an input for the degradation element for ctnnb1. Erk influences the downstream transcipriton factors by transport out of the nucleus. This export is simulated by the Michaelis Menten element. The core network was built of OR-logic combined with repressors. The Oct enhancer was designed according to [29] or the cooperative element with repression respectively. All genes in the Feedback and differentiation layer were set up as Or-logic elements combined with repressors.

2.3.3 Values assumed in the simulation

From measurements [41] we estimated ranges of orders of magnitude of transcription and translation rates and mRNA and Protein half-lifes of

$$\begin{aligned} \nu_{Transcription} &= [1e-5, 5e-1] \cdot s^{-1} \\ \nu_{Translation} &= [1e-2, 10] \cdot s^{-1} \cdot mRNA^{-1} \\ \delta_{mRNA} &= [1e-5, 1e-2] s^{-1} \\ \delta_{\mathrm{Protein}} &= [1\text{e-}3, 1e-9] s^{-1} \end{aligned} \tag{1.94}$$

Since we did not know the exact values for transcription and translation, we used the values above to conclude effective birth and death rates for the simulation. These effective rates correspond to a sequence of transcription of a gene with rate $\nu_{Transcription}$ and translation of the transcript with rate $\nu_{Translation}$. Depending on the mRNA levels these rates may be in the range of

$$\begin{aligned} \nu &= [1e-7, 1] \cdot s^{-1} \\ \delta &= [1\text{e-}2, 1e-9] s^{-1} \end{aligned} \tag{1.95}$$

For the on-rates of genes we estimated the diffusion limited on-rate in a nucleus also taking into account observed target search times

$$\lambda = [1e-5, 1e-2] \cdot s^{-1} \tag{1.96}$$

For off-rates we took measurements of TF-binding times and estimated a range of binding times relevant for transcription regulation

$$\mu = [1e-3, 1] \cdot s^{-1} \tag{1.97}$$

Where possible, we used experimentally observed values from literature. We designed the Sox2/Oct4 enhancer including binding times of Oct4 and Sox2 as reported in [29]. In addition, we applied the binding times of Stat3 and Klf4[42]. Degradation rates for the core pluripotency factors Klf4, Nanog, Sox2 and Oct4 measured in [43] were also included in the network.

Due to missing information of specific rate constants of promoters, mRNA and proteins we guessed standard parameter-sets depending on the function that we ascribe to the respective promoter or protein.

For static expression levels, e.g. transcription factors which are always present, but whose activity is controlled by phosphorylation we took the parameter-set

$$\text{Self Sustained Expression: } \lambda = 1e-2 \quad \mu = 1e-2 \quad \nu = 1e-2 \quad \delta = 1e-5 \tag{1.98}$$

where the on rate λ is not connected to any TF-species in the network. For enzyme dynamics, we assumed that enzyme-substrate-complex formation is limited by diffusion. Product synthesis rates and affinities were chosen such that the enzyme performs its biological function.

Supplementary Table 4: Parameters for the pluripotency network. Biomolecules with the suffix "-I" represent inhibited versions.

Species	Interaction	Rate	Birth/Death	Ref.
Stat3Dimer	Lif	$\lambda_1 = 1e-3; \mu_1 = 1; \gamma_1 = 1$	$\gamma_2 = 1$	11
	Stat3Dimer	$\lambda_3 = 1e-3; \mu_3 = 1$		
	Socs	$\lambda_2 = 1e-3; \mu_2 = 1e-5$		
Stat3	Self-Sustained Expression Parameter-Set			(1.98)
Socs	Stat3Dimer	$\lambda = 1e-3; \mu = 1e-1$	$\nu = 1e-2; \delta = 1e-4$	11
Klf4	Stat3Dimer	$\lambda = 1e-3; \mu = 12e-2$	$\nu = 3.5e-3; \delta = 3.5e-5$	12
Klf4	Erk	$\lambda = 1e-3; \mu = 1; \nu = 1; \gamma = 1$	n/a	20
Smad2/3	Self-Sustained Expression Parameter-Set			(1.98)
Smad23#	TGF-b	$\lambda = 1e-3; \mu = 1; \nu = 1; \gamma = 1$	n/a	16
Smad158	Self-Sustained Expression Parameter-Set			(1.98)
Smad158#	BMP	$\lambda = 1e-3; \mu = 1; \nu = 1; \gamma = 1$	n/a	17
Dusp	Smad158#	$\lambda = 1e-3; \mu = 1e-1$	$\nu = 1e-1; \delta = 1e-3$	17
Erk-I	Self-Sustained Expression Parameter-Set		$\nu = 1e-2; \delta = 5e-5$	-
Erk	Fgf+Lif	$\lambda_1 = 1e-5; \mu_1 = 0.1; \gamma_1 = 0.1$	n/a	19
	Dusp	$\lambda_2 = 1e-5; \mu_2 = 0.1; \gamma_2 = 10$		6
GSK	Self-Sustained Expression Parameter-Set			-
GSK	Wnt	$\lambda = 1e-3; \mu = 1; \nu = 1; \gamma = 1e-2$	n/a	6
Ctnnb1	Self-Sustained Expression Parameter-Set		$\nu = 1e-1; \delta = 1e-3$	-
Ctnnb1	GSK#	$\lambda = 1e-3; \mu = 1; \gamma = 0.01$		6
Tcf3-I	Self-Sustained Expression Parameter-Set		$\nu = 1e-2; \delta = 1e-4$	-
	Ctnnb1	$\lambda = 1e-3; \mu = 1e-3$		23
	ID	$\lambda = 1e-3; \mu = 1e-3$		18
Tcf	Erk	$\lambda = 1e-3; \mu = 1; \nu = 1e-1; \gamma = 1e-4$	n/a	21
ID	Smad158#	$\lambda = 1e-3; \mu = 1e-1$	$\nu = 1e-1; \delta = 1e-3$	6
Nanog	Nanog	$\lambda = 1e-3; \mu = 1e1$	$\nu = 1.4e-2; \delta = 1.4e-4$	6
	Klf4exp	$\lambda = 1e-3; \mu = 1e-1$		14
	Oct4	$\lambda = 1e-3; \mu = 68e-3$		27
	Smad23	$\lambda = 1e-3; \mu = 1e-1$		16
	Tcf3Exp	$\lambda = 1e-3; \mu = 1e-1$		24
Sox2	Nanog	$\lambda = 1e-3; \mu = 1e-1$	$\nu = 3.5e-3; \delta = 3.5e-5$	27
	Sox2	$\lambda = 1e-3; \mu = 8e-2$		28
Oct4	Sox2	$\lambda = 1e-3; \mu = 8e-2$	$\nu = 7e-3; \delta = 7e-5$	27
	Oct4	$\lambda = 1e-3; \mu = 68e-3$		44
	Sall4	$\lambda = 1e-3; \mu = 1$		37

	Nr2f2	$\lambda = 1e-3; \mu = 1e-2$		25
Mir302	Oct4	$\lambda = 1e-3; \mu = 68e-3$	$\nu = 1e-1; \delta = 1e-3$	23
	TCF3Exp	$\lambda = 1e-3; \mu = 1e-3$		23
Nr2f2	Self-Sustained Expression Parameter-Set			
	Mir302a	$\lambda = 1e-3; \mu = 1e-3$	$\nu = 1e-2; \delta = 1e-4$	26
Sall4	Oct4	$\lambda = 1e-3; \mu = 68e-3$	$\nu = 5e-2; \delta = 1e-3$	37
	Sall4	$\lambda = 1e-3; \mu = 1$		37
Six3	Sox2	$\lambda = 1e-3; \mu = 8e-2$	$\nu = 1e-2; \delta = 1e-4$	34
	Pax6	$\lambda = 1e-3; \mu = 5e-2$		36
Pax6	Six3	$\lambda = 1e-4; \mu = 1e-2$	$\nu = 1e-2; \delta = 1e-4$	36
	Pax2	$\lambda = 1e-3; \mu = 1e-3$		40
Pax2	Oct4	$\lambda = 1e-3; \mu = 68e-3$	$\nu = 1e-1; \delta = 1e-3$	39
	Pax6	$\lambda = 1e-3; \mu = 1e-3$		40
Gata6	Self-Sustained Expression Parameter-Set		$\nu = 1e-2; \delta = 1e-4$	-
Gata6	Basal Expr.	$\lambda = 1e-3; \mu = 1e-3$	$\nu = 1e-2; \delta = 1e-4$	-
	Gata6#	$\lambda = 1e-3; \mu = 1e-1$		22
	Sall4	$\lambda = 1e-3; \mu = 1e-3$		38
Gata6#	Erk	$\lambda = 1e-3; \mu = 1e-1; \gamma_1 = 1; \gamma_2 = 1e-2$		22
Snai1	Nanog	$\lambda = 1e-3; \mu = 1e-1$	$\nu = 1e-1; \delta = 1e-3$	32
	Snai1	$\lambda = 1e-3; \mu = 1$		33
Runx2	Smad158#	$\lambda = 1e-3; \mu = 1e-1$	$\nu = 1e-1; \delta = 1e-3$	17
	Snai1	$\lambda = 1e-3; \mu = 1e-3$		33

2.3.4 Culture condition-dependent differentiation

To model transition of the pluripotency gene regulatory network from the Naïve state over the primed state into the three germ lines, we considered the following input signal patterns. Naïve pluripotency is maintained by Lif and BMP signaling [45]. Primed pluripotency is initiated and maintained by TGFB and FGF [45]. Ectoderm differentiation is the default pathway if the core network is shut down without any further inputs [46]. Mesoderm differentiation is supported by BMP [47]. Endoderm differentiation is supported by Erk [22], which is activated by Fgf signaling.

2.3.5 Design of a differentiation element

Here we investigate a recurring motif associated with differentiation in the pluripotency network. It is composed of a repressor that is activated by pluripotency factors and inhibits transcription of a differentiation factor. The differentiation factor is activated by an extracellular signaling molecule associated with a certain differentiation. An example is Snail1, which is activated by Nanog and represses Runx2, while Runx2 is activated by Smad1/5/8, a signal transduction factor activated by Bmp.

For the motif to be functional, repression of the differentiation marker by the pluripotency factor and subsequent repressor needs to dominate activation of the differentiation marker by the differentiation signal. Differentiation than only occurs if the core pluripotency network is shut down. All combinations of inputs and the desired outputs of the motif are summarized in Supplementary Table 5.

Supplementary Table 5: Target functionality of the differentiation motif. Independent of the Differentiation signal, the pluripotency signal has to repress the differentiation marker.

Pluripotency Signal	Differentiation Signal	Output Differentiation Marker
Low	Low	Low
Low	High	High
High	Low	Low
High	high	low

To identify kinetic parameters, in particular the on- and off-rates of genes, we have designed a target function to combine all different outputs and assigned a score to the Parameters.

To design the target function, we first quantify the fuzzy terms “low” and “high”. We define thresholds with the following functions

$$\begin{aligned} T_{active}(x)&=\text{atan}(6(x-0.6))/\text{pi}+0.5;\\ T_{inactive}(x)&=1-(\text{atan}(6(x-0.3))\text{pi}+0.5); \end{aligned} \tag{1.99}$$

We then use these thresholds to combine all conditions by multiplying the thresholds of the corresponding conditions resulting in the cost-function

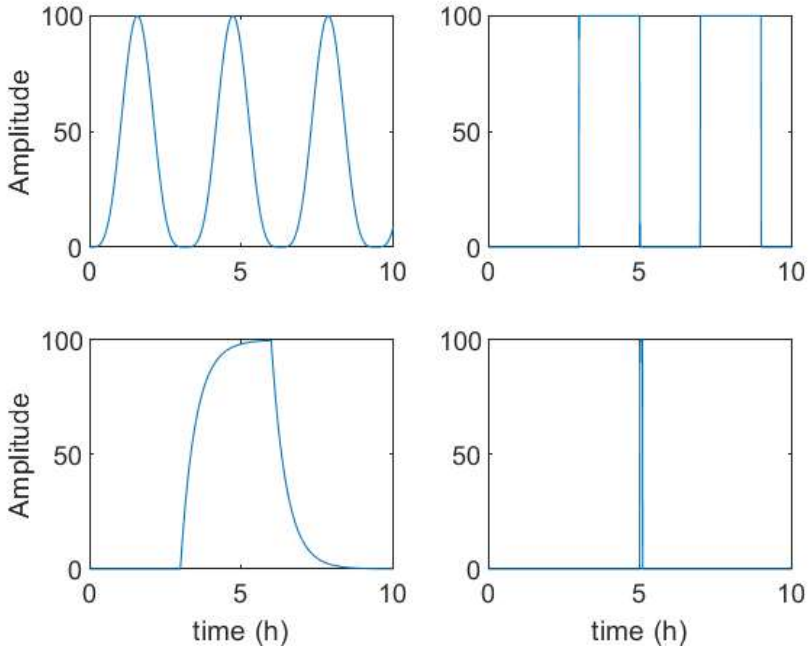
$$s=\prod_{Conditions active} T_{active}(x) \prod_{Condition \sin active} T_{inactive}(x) \tag{1.100}$$

We set up the network motif depicted in (Figure 3 panel j) in CaiNet using two input elements for pluripotency signals and Differentiation signals. We implemented the differentiation marker (here Runx2) as an OR-logic and the Repressor with an or logic. We then applied all pulse combinations (Table 1) to this small network. To test whether a specific parameter set is required for the motif to work we varied the binding times of the Differentiation and the Pluripotency factor. We found that the cost-function has a distinct minimum for a single set or parameters.

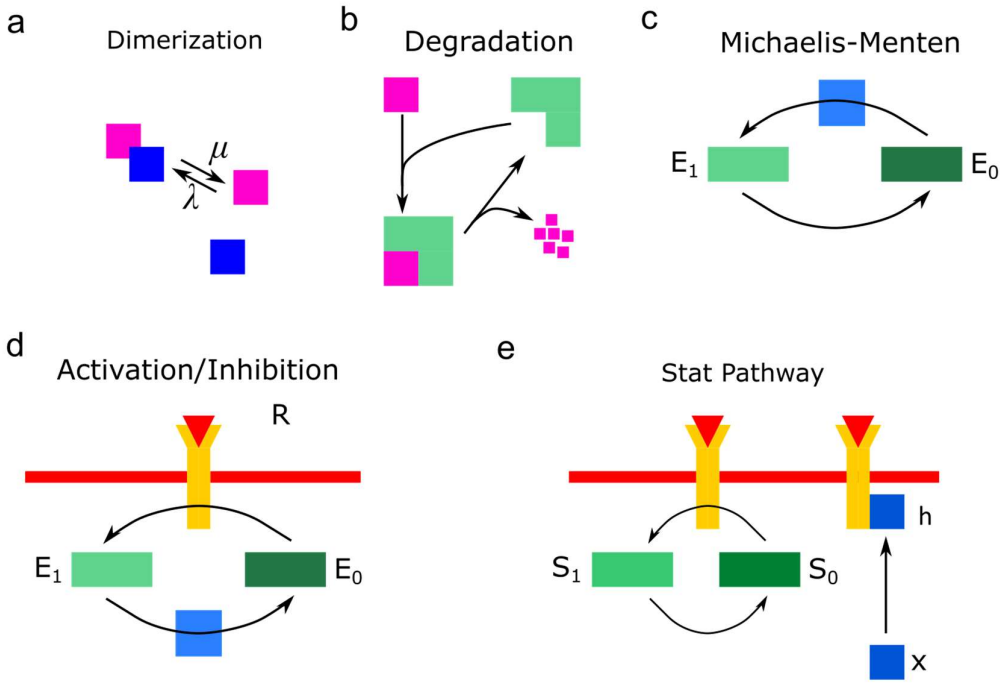
References

1. Sanft, K.R. et al. StochKit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics* **27**, 2457-2458 (2011).
2. Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20**, 3289-3291 (2004).
3. Brown, C.T. et al. New computational approaches for analysis of cis-regulatory networks. *Developmental Biology* **246**, 86-102 (2002).
4. Funahashi, A. et al. CellDesigner 3.5: A versatile modeling tool for biochemical networks. *Proceedings of the Ieee* **96**, 1254-1265 (2008).
5. Slepoy, A., Thompson, A.P. & Plimpton, S.J. A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. *Journal of Chemical Physics* **128** (2008).
6. Ogata, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **27**, 29-34 (1999).
7. Welch, R.D. & Flaveny, C.A. REV-ERB and ROR: therapeutic targets for treating myopathies. *Physical Biology* **14** (2017).
8. Xiao, J. et al. Transcription Factor NF-Y Is a Functional Regulator of the Transcription of Core Clock Gene Bmal1. *Journal of Biological Chemistry* **288**, 31930-31936 (2013).
9. Han, H. et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research* **46**, D380-D386 (2018).
10. Papatsenko, D., Waghray, A. & Lemischka, I.R. Feedback control of pluripotency in embryonic stem cells: Signaling, transcription and epigenetics. *Stem Cell Research* **29**, 180-188 (2018).
11. Kile, B.T. et al. The SOCS box: a tale of destruction and degradation. *Trends in Biochemical Sciences* **27**, 235-241 (2002).
12. Hall, J. et al. Oct4 and LIF/Stat3 Additively Induce Kruppel Factors to Sustain Embryonic Stem Cell Self-Renewal. *Cell Stem Cell* **5**, 597-609 (2009).
13. Guerriero, M.L., Dudka, A., Underhill-Day, N., Heath, J.K. & Priami, C. Narrative-based computational modelling of the Gp130/JAK/STAT signalling pathway. *Bmc Systems Biology* **3**, 12 (2009).
14. Zhang, P.L., Andrianakos, R., Yang, Y., Liu, C.M. & Lu, W.G. Kruppel-like Factor 4 (Klf4) Prevents Embryonic Stem (ES) Cell Differentiation by Regulating Nanog Gene Expression. *Journal of Biological Chemistry* **285**, 9180-9189 (2010).
15. Jo, C. et al. Leukemia inhibitory factor blocks early differentiation of skeletal muscle cells by activating ERK. *Biochimica Et Biophysica Acta-Molecular Cell Research* **1743**, 187-197 (2005).
16. Sakaki-Yumoto, M., Liu, J.M., Ramalho-Santos, M., Yoshida, N. & Derynck, R. Smad2 Is Essential for Maintenance of the Human and Mouse Primed Pluripotent Stem Cell State. *Journal of Biological Chemistry* **288**, 18546-18560 (2013).
17. Liu, Z.Y. et al. Molecules mimicking Smad1 interacting with hox stimulate bone formation. *Journal of Biological Chemistry* **279**, 11313-11319 (2004).
18. Rohde, M. et al. Relevance of ID3-TCF3-CCND3 pathway mutations in pediatric aggressive B-cell lymphoma treated according to the non-Hodgkin Lymphoma Berlin-Frankfurt-Munster protocols. *Haematologica* **102**, 1091-1098 (2017).
19. Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* **137**, 3351-3360 (2010).
20. Dhaliwal, N.K., Miri, K., Davidson, S., El Jarkass, H.T. & Mitchell, J.A. KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency. *Stem Cell Reports* **10**, 1308-1323 (2018).
21. Shan, J. et al. Tcf7l1 acts as a suppressor for the self-renewal of liver cancer stem cells and is regulated by IGF/MEK/ERK signaling independent of β-catenin. *STEM CELLS* (2019).
22. Meng, Y. et al. GATA6 phosphorylation by Erk1/2 propels exit from pluripotency and commitment to primitive endoderm. *Developmental Biology* **436**, 55-65 (2018).
23. Brautigam, C., Raggioli, A. & Winter, J. The Wnt/beta-Catenin Pathway Regulates the Expression of the miR-302 Cluster in Mouse ESCs and P19 Cells. *Plos One* **8** (2013).
24. Yi, F., Pereira, L. & Merrill, B.J. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* **26**, 1951-1960 (2008).
25. Benshushan, E., Sharir, H., Pikarsky, E. & Bergman, Y. A DYNAMIC BALANCE BETWEEN ARP-1/COUP-TFII, EAR-3/COUP-TFI, AND RETINOIC ACID RECEPTOR RETINOID-X RECEPTOR HETERODIMERS REGULATES OCT-3/4 EXPRESSION IN EMBRYONAL CARCINOMA-CELLS. *Molecular and Cellular Biology* **15**, 1034-1048 (1995).
26. Sinkkonen, L. et al. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nature Structural & Molecular Biology* **15**, 259-267 (2008).
27. Dunn, S.J., Martello, G., Yordanov, B., Emmott, S. & Smith, A.G. Defining an essential transcription factor program for naive pluripotency. *Science* **344**, 1156-1160 (2014).
28. Strebinger, D. et al. Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions. *Molecular Systems Biology* **15**, e9002 (2019).
29. Chen, J.J. et al. Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell* **156**, 1274-1285 (2014).

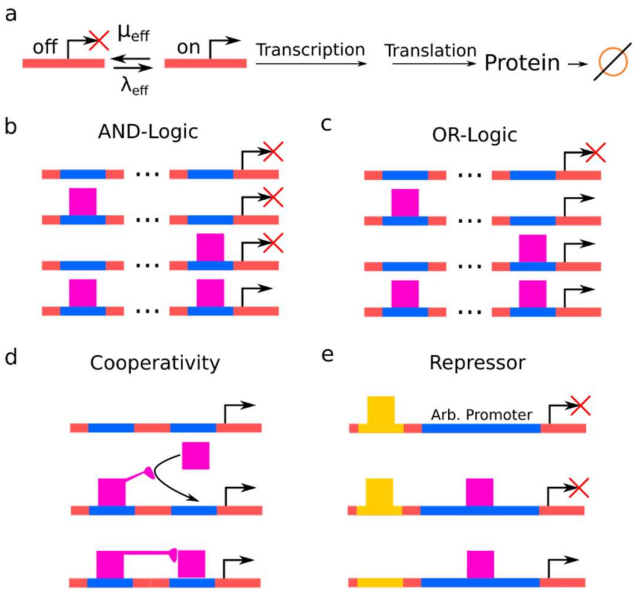
30. Wei, J.W. et al. Glucose Uptake and Runx2 Synergize to Orchestrate Osteoblast Differentiation and Bone Formation. *Cell* **161**, 1576-1591 (2015).
31. Osumi, N., Shinohara, H., Numayama-Tsuruta, K. & Maekawaa, M. Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. *Stem Cells* **26**, 1663-1672 (2008).
32. Gingold, J.A. et al. A Genome-wide RNAi Screen Identifies Opposing Functions of Snai1 and Snai2 on the Nanog Dependency in Reprogramming. *Molecular Cell* **56**, 140-152 (2014).
33. Park, S.J. et al. The transcription factor snail regulates osteogenic differentiation by repressing Runx2 expression. *Bone* **46**, 1498-1507 (2010).
34. Lee, B. et al. Genomic code for Sox2 binding uncovers its regulatory role in Six3 activation in the forebrain. *Developmental Biology* **381**, 491-501 (2013).
35. Marsich, E., Vetere, A., Di Piazza, M., Tell, G. & Paoletti, S. The PAX6 gene is activated by the basic helix-loop-helix transcription factor NeuroD/BETA2. *Biochemical Journal* **376**, 707-715 (2003).
36. Goudreau, G. et al. Mutually regulated expression of Pax6 and SW and its implications for the Pax6 haploinsufficient lens phenotype. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8719-8724 (2002).
37. Yang, J.C., Gao, C., Chai, L. & Ma, Y.P. A Novel SALL4/OCT4 Transcriptional Feedback Network for Pluripotency of Embryonic Stem Cells. *Plos One* **5** (2010).
38. Lim, C.Y. et al. Sall4 Regulates Distinct Transcription Circuitries in Different Blastocyst-Derived Stem Cell Lineages. *Cell Stem Cell* **3**, 543-554 (2008).
39. Ramos-Mejia, V. et al. Phenotypic analyses of mouse embryos with ubiquitous expression of Oct4: Effects on mid-hindbrain patterning and gene expression. *Developmental Dynamics* **232**, 180-190 (2005).
40. Schwarz, M. et al. Spatial specification of mammalian eye territories by reciprocal transcriptional repression of Pax2 and Pax6. *Development* **127**, 4325-4334 (2000).
41. Fuchs, G. et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biology* **15** (2014).
42. Xie, L.Q. et al. A dynamic interplay of enhancer elements regulates Klf4 expression in naive pluripotency. *Genes & Development* **31**, 1795-1808 (2017).
43. Alber, A.B., Paquet, E.R., Biserni, M., Naef, F. & Suter, D.M. Single Live Cell Monitoring of Protein Turnover Reveals Intercellular Variability and Cell-Cycle Dependence of Degradation Rates. *Molecular Cell* **71**, 1079-+ (2018).
44. Chew, J.L. et al. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Molecular and Cellular Biology* **25**, 6031-6046 (2005).
45. Morgani, S., Nichols, J. & Hadjantonakis, A.K. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. *Bmc Developmental Biology* **17** (2017).
46. Munoz-Sanjuan, I. & Brivanlou, A.H. Neural induction, the default model and embryonic stem cells. *Nature Reviews Neuroscience* **3**, 271-280 (2002).
47. Loh, K.M. et al. Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451-467 (2016).



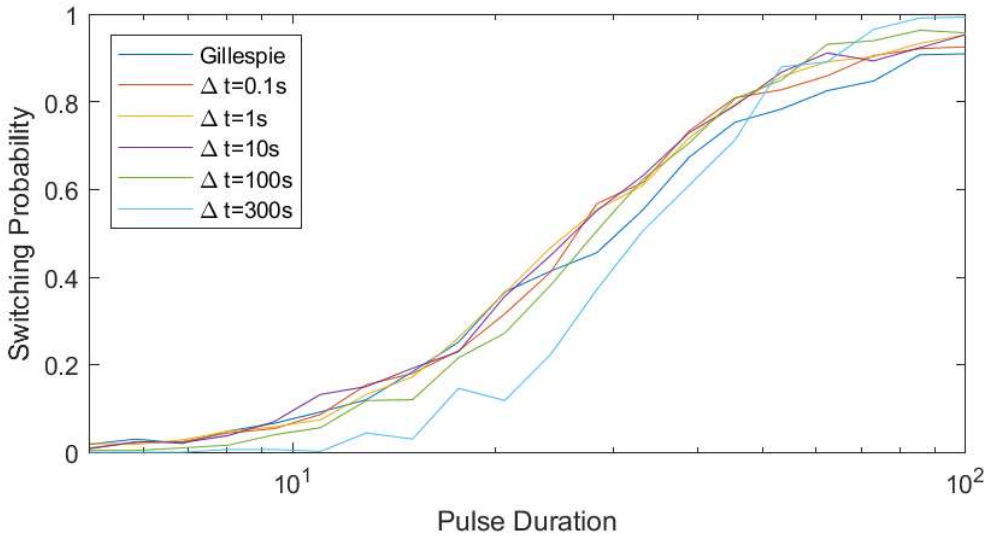
Supplementary Figure 1: Examples for possible input trajectories. Sinusoidal (a), pulsed plateaus (b), smoothed plateaus (c) and sharp pulses (d).



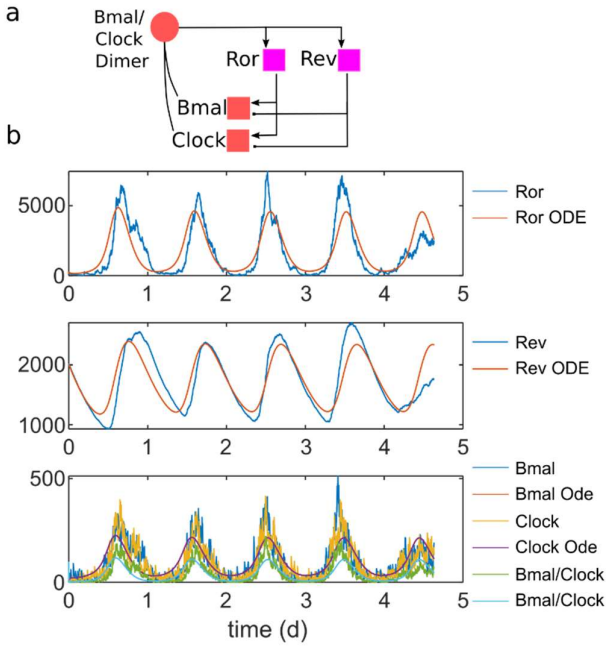
Supplementary Figure 2: Biochemical Reactions. (a) Two monomers of different or same species can form a dimer (b) The degradation of species (purple) s controlled by an enzyme (green). (c) The reaction of the reagent E_0 to E_1 is catalyzed by an enzyme (blue). The reverse reaction is not catalyzed. (d) Both, the forward and reverse reaction of E_0 to E_1 is catalyzed by two different enzymes. (e) The reaction in (c) is enhanced by an inhibitor that deactivates the Enzymes.



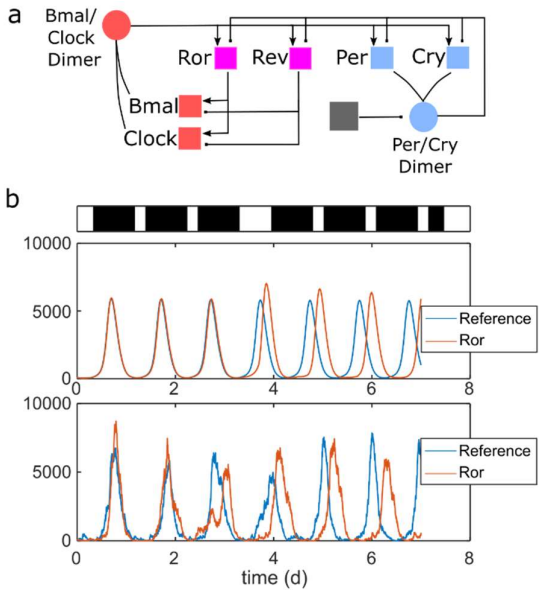
Supplementary Figure 3: Promoter Structures. (a) We approximate all promoter with a two-state model, where a gene can either produce mRNA or remain silent. Until the Gene-Product is finished delay steps due to transcription and translation have to be passed. (b) AND-Logic for promoter activation. The Promoter is only active if all activating transcription factors are bound. (c) OR-logic for promoter activation. The promoter is active as soon as a single activating TF is bound. (d) A repressor can be added to all the aforementioned promoter structures. If the Repressor is bound, transcription is repressed independent of the state of the promoter.



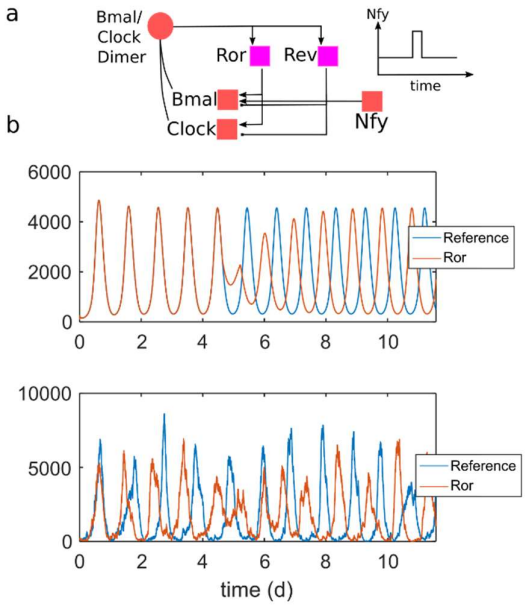
Supplementary Figure 4: Influence of Time steps. Up to a time step of 100 the pulse-duration of CaiNet compares well to the result of the exact Gillespie-Simulation. For $\Delta t = 300$ CaiNet differs from the exact result of the Gillespie algorithm.



Supplementary Figure 5: Internal Loop of the circadian clock. (a) Circuitry of the internal loop of the circadian clock. (b) During a period of 4 days the stochastic internal oscillation of the internal loop is coherent with a deterministic 24-h oscillation.



Supplementary Figure 6: Phase Shift induced by a pulsed shift in Per expression. (a) Circuitry of the internal loop complemented with Per. (b) a pulse of active Per causes repression of the internal loop and forces a phase shift.



Supplementary Figure 7:(a) Scheme of spontaneous activation of the Bmal/Clock promoter. (b) If the level of NFY is modulated for a short time-period the oscillation experiences a phase shift.