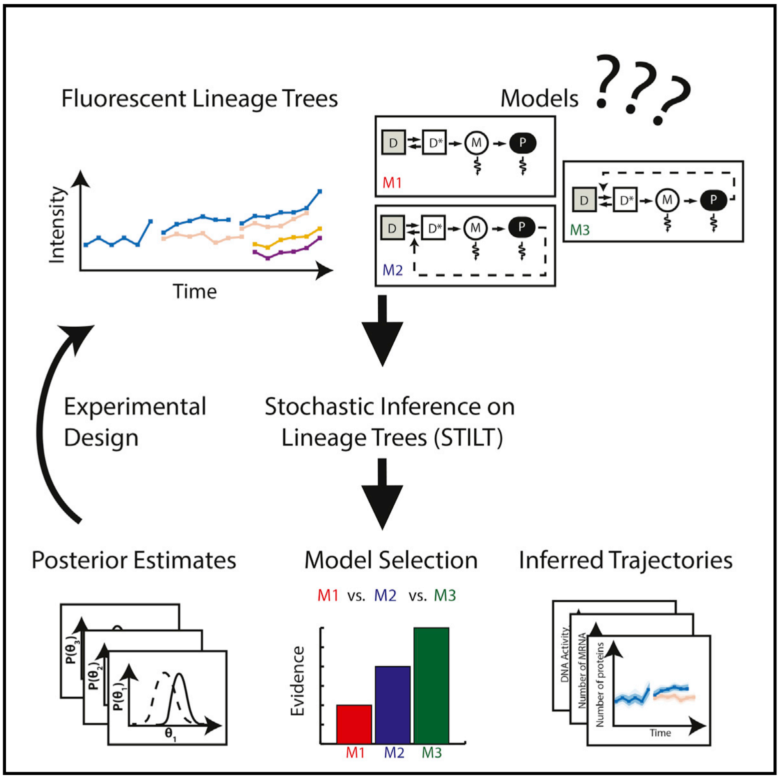
Cell Systems

Analysis of Cell Lineage Trees by Exact Bayesian Inference Identifies Negative Autoregulation of Nanog in Mouse Embryonic Stem Cells

Graphical Abstract



Authors

Justin Feigelman, Stefan Ganscha,
Simon Hastreiter, ..., Fabian J. Theis,
Carsten Marr, Manfred Claassen

Correspondence

carsten.marr@helmholtz-muenchen.de
(C.M.),
manfred.claassen@imsb.biol.ethz.ch
(M.C.)

In Brief

Using a novel Bayesian inference method developed for fluorescence lineage trees, we compare possible models of Nanog autoregulation in mouse embryonic stem cells. We identify negative feedback as a likely mode of Nanog regulation and verify this using a novel fluorescence dataset.

Highlights

- We present a method for fitting stochastic models to single-cell fluorescent lineage trees
- We analyze Nanog expression obtained from populations of mouse embryonic stem cells
- We evaluate autoregulation models and find poor agreement for positive feedback
- We rationally design overexpression experiments and identify negative feedback

Feigelman et al., 2016, Cell Systems *3*, 480–490
November 23, 2016 © 2016 The Authors. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.cels.2016.11.001

OPEN
ACCESS
CellPress

Cell Systems
Report

Analysis of Cell Lineage Trees by Exact Bayesian Inference Identifies Negative Autoregulation of Nanog in Mouse Embryonic Stem Cells

Justin Feigelman,[1,2,5] Stefan Ganscha,[1] Simon Hastreiter,[3] Michael Schwarzfischer,[2] Adam Filipczyk,[4] Timm Schroeder,[3] Fabian J. Theis,[2,5] Carsten Marr,[2,*] and Manfred Claassen[1,6,*]
[1]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany
[2]Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland
[3]Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland
[4]Department of Microbiology, Oslo University Hospital, 0450 Oslo, Norway
[5]Department of Mathematics, Technische Universität München, 85748 Garching, Germany
[6]Lead Contact
*Correspondence: carsten.marr@helmholtz-muenchen.de (C.M.), manfred.claassen@imsb.biol.ethz.ch (M.C.)
http://dx.doi.org/10.1016/j.cels.2016.11.001

SUMMARY

Many cellular effectors of pluripotency are dynamically regulated. In principle, regulatory mechanisms can be inferred from single-cell observations of effector activity across time. However, rigorous inference techniques suitable for noisy, incomplete, and heterogeneous data are lacking. Here, we introduce stochastic inference on lineage trees (STILT), an algorithm capable of identifying stochastic models that accurately describe the quantitative behavior of cell fate markers observed using time-lapse microscopy data collected from proliferating cell populations. STILT performs exact Bayesian parameter inference and stochastic model selection using a particle-filter-based algorithm. We use STILT to investigate the autoregulation of Nanog, a heterogeneously expressed core pluripotency factor, in mouse embryonic stem cells. STILT rejects the possibility of positive Nanog autoregulation with high confidence; instead, model predictions indicate weak negative feedback. We use STILT for rational experimental design and validate model predictions using novel experimental data. STILT is available for download as an open source framework from http://www.imsb.ethz.ch/research/claassen/Software/stilt---stochastic-inference-on-lineage-trees.html.

INTRODUCTION

Time-lapse fluorescence microscopy provides a means to unambiguously label, track, and quantify transcriptional and translational expression dynamics of individual cells, providing insight into the dynamics of gene expression (Coutu and Schroeder, 2013; Elowitz et al., 2002; Larson et al., 2011; Muzzey and van Oudenaarden, 2009; Raj et al., 2006). Using time-lapse fluorescence microscopy, it is possible to quantitatively monitor expression dynamics, such as the onset of lineage-determining markers in progenitor cells over the course of days (Eilken et al., 2009; Hoppe et al., 2016; Rieger et al., 2009) or heterogeneous response to external perturbation (Spencer et al., 2009; Tay et al., 2010) at the transcript and protein level over the course of minutes to hours. Movies obtained from time-lapse microscopy provide valuable dynamic intra- and intercellular correlations, which are informative of the underlying regulatory mechanisms (Muzzey and van Oudenaarden, 2009) and thus well suited for the investigation of causal relationships between genes (Dunlop et al., 2008; Locke and Elowitz, 2009). Moreover, specialized tools, such as TtT/QTFy (Hilsenbeck et al., 2016), for tracking, segmenting, and quantifying time-lapse fluorescence microscopy movies facilitate the wide adoption of such a strategy.

Despite its power, however, time-lapse fluorescence microscopy presents a number of obstacles that must be overcome in order to fully exploit the available data. Measurements are often noisy and incomplete, with only one or a few of the potentially interesting species interrogable via fluorescence imaging due to challenges in the construction of reporter systems. They may also be obtained infrequently, yielding uncertainty about transitions between observations. Besides measurement error, the studied systems also typically contain significant biological noise arising from biological variation and inherent stochasticity (Raj and van Oudenaarden, 2008). For example, many transcripts are present with low copy number, which can give rise to substantial intercellular variation in expression (Becskei et al., 2005; Schwanhäusser et al., 2011).

One approach to address such challenges is to employ stochastic models of gene regulation (Wilkinson, 2009). Stochastic models capture cellular variability arising from intrinsic noise and are thus superior to deterministic methods, e.g., ordinary differential equations, for modeling such data. In particular, stochastic "chemical reaction network" models (see Box 1, top, for glossary of italicized terms) reflect the probabilistic evolution of a set of chemical species via reactions with probabilities that vary according to "mass action kinetics". Such models can be fit to time-lapse data using "Bayesian inference" approaches, in particular "particle filters" (Doucet and Johansen, 2009),

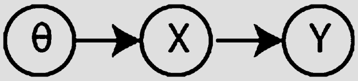
Box 1. Glossary Terms and STILT

GLOSSARY OF TERMS

Chemical reaction network: a system of chemical species that interact via a collection of defined reactions and associated kinetic constants
Mass action kinetics: a kinetic model where the reaction rate is proportional to the quantity of each educt exponentiated to its stoichiometric coefficient
Bayesian inference: an inference procedure incorporating both likelihood of observation and prior knowledge of the probability of parameter values
Particle: a sample of model parameters and/or stochastic trajectories
Particle filter: an inference method to approximate the true posterior distribution of model parameters and/or trajectories using a mixture of sampled particles
Latent species: a species in the chemical reaction network that is not directly observed
Latent history: the unobserved trajectories of a set of latent species on a lineage tree
Evidence: the total probability of a model integrated over all possible parameter values and stochastic trajectories
Prior distribution: the probability distribution of a parameter based on prior experimental or literature knowledge
Bayes factor: the ratio of evidences between two models
Posterior distribution: the probability distribution of a parameter based on both the prior distribution and the likelihood of new experimental observations
Goodness-of-fit test: a test to decide whether a model (and parameters) is compatible with observed data
Subtree: a lineage tree formed by extracting a group of cells from a larger lineage tree

STILT

The aim of STILT is to recover the latent history of species $\boldsymbol{X}$ (here, DNA, mRNA, and protein time series) in the chemical reaction network, for all cells, and the set of model parameters θ (here, the reaction constants specifying the chemical reaction network). The latent history $\boldsymbol{X}$ depends on the reaction constants θ, and the observations of the latent history, $\boldsymbol{Y}$, depend on $\boldsymbol{X}$.



The STILT algorithm consists of the following steps:

1. Specify a chemical reaction network of interest (species and reaction stoichiometries)
2. Specify prior distributions for the initial state $\boldsymbol{X}$ and model parameters $\boldsymbol{\theta}$
3. Sample a collection of P particles. Each particle consists of a set of parameters $\boldsymbol{\theta}^k$ and state $\boldsymbol{X}^k$, drawn independently from their prior distributions, $k = 1, \ldots, P$
4. For each measurement time point t_i
    - 4.1. Stochastically simulate each particle k according to the assumed chemical reaction network, conditional on $(\boldsymbol{X}^k, \boldsymbol{\theta}^k)$, to obtain a new state $\boldsymbol{X}'^k$
    - 4.2. Evaluate the likelihood of each particle $w_i^k = P(\boldsymbol{Y}_{t_i} | \boldsymbol{X}'^k)$
    - 4.3. Generate a new set of particles $\boldsymbol{X}^k$ by sampling (with replacement) from the collection $\boldsymbol{X}'^k$ proportionally to the likelihoods w_i^k
    - 4.4. Sample new parameters $\boldsymbol{\theta}^k \sim P(\boldsymbol{\theta} | \boldsymbol{X}^k)$

The result of each iteration of the algorithm is a set of particles sampled from the posterior distribution $P(\boldsymbol{X}, \boldsymbol{\theta} | \boldsymbol{Y})$ using the observations until that time point. STILT also estimates the marginal evidence of the model P(**Y**), which is useful for model selection. STILT extends the basic bootstrap particle filter algorithm by incorporating cell division events in the forward simulation. When cells divide, the contents of the mother cell are allocated randomly to the daughter cells. After "birth," each cell is simulated independently in step 4.1, and the likelihoods of the composite simulated trajectory are multiplied to achieve the total likelihood of that particle in step 4.2. Particles are then resampled according to the composite likelihood in step 4.3.

which attempt to learn the unknown state of the system (e.g., of unobserved species) and biochemical reaction constants. Inference results take the form of posterior probability distributions over the parameters and trajectories or "latent histories" of all modeled chemical species. Although particle-filter-based approaches have been successfully applied to the inference of single-cell trajectories (Zechner et al., 2011), inference for populations of proliferating cells presents a formidable challenge as the number of "latent species" and thus the problem size grows at each cellular division event.

Ideally, one could exploit the cellular genealogy once it is known in order to take advantage of the additional constraints that it imposes upon the allowable state of the system. For example, upon division, mass is conserved such that the daughter cells' inherited cellular contents equal that of the mother cell. Additionally, any proposed parameters must be likely in the context of all cells simultaneously and not just each cell individually, which constrains the set of plausible model parameters. However, so far, no computational algorithm has been provided for Bayesian inference of latent histories and parameter values on tree-structured data. In this work, we extend the recursive particle filter (Pitt and Shephard, 1999) to explicitly model cell division, thus enabling inference of stochastic gene-regulatory models using data obtained from long-term fluorescent time-lapse microscopy of cellular lineage trees. The proposed algorithm, stochastic inference on lineage trees (STILT) (Box 1, bottom), is suitable for proliferating single-cell data that are noisy, partially and discretely observed, heterogeneous, and with small molecule numbers and intrinsic stochasticity. STILT is available for download as an open source framework for gene-regulation models specified using SBML (Hucka et al., 2003) from http://www.imsb.ethz.ch/research/claassen/Software/stilt---stochastic-inference-on-lineage-trees.html.

Using STILT, we study the stochastic expression dynamics of Nanog in mouse embryonic stem cells (mESCs). Nanog is a key regulator of pluripotency, whose expression is fundamentally stochastic, involving the chance synthesis, degradation, and interaction of biochemical species. It is heterogeneously expressed (Chambers et al., 2007; Filipczyk et al., 2015; Singer et al., 2014) and exhibits strong fluctuations in expression (Abranches et al., 2014; Filipczyk et al., 2015), which may serve to prime mESCs for differentiation (Abranches et al., 2014; Torres-Padilla and Chambers, 2014). Nanog forms homodimers (Mullin et al., 2008; Wang et al., 2008) and binds its own enhancer (Saunders et al., 2013), and Nanog-dependent feedback loops are thought to be critical to mESC regulation (MacArthur et al., 2012). However, Nanog's mode of autoregulation has been debated. Whereas Nanog has long been thought to exhibit positive autoregulation (Boyer et al., 2005; Jaenisch and Young, 2008), recent studies have provided evidence for both negative feedback (Fidalgo et al., 2012; Navarro et al., 2012) and no direct feedback (Ochiai et al., 2014). To ultimately discriminate between such mechanisms, quantitative model selection based on single-cell data is required.

Several recent investigations have provided a view of expression dynamics of Nanog in mESCs at the single cell level (Abranches et al., 2013; Filipczyk et al., 2015; Ochiai et al., 2014; Singer et al., 2014). However, to our knowledge, no attempt has been made until now to directly infer stochastic dynamic models from Nanog time-lapse data, infer parameters, or perform model selection for competing regulatory motifs, owing in part to the computational challenges presented by such stochastic, tree-structured data. We use STILT to investigate candidate models for Nanog autoregulation directly from time-lapse fluorescence data. We assess the models' overall likelihood using a collection of previously published fluorescence lineage trees (Filipczyk et al., 2015), infer model parameters, and determine the compatibility of each model with the inferred parameters for the data analyzed. By computing model "evidences," we determine positive feedback to be an unlikely mode of Nanog autoregulation. To resolve between no feedback and negative feedback, we utilize STILT to rationally design an informative perturbation experiment and predict and subsequently verify its response, finally identifying weak negative feedback as Nanog's most probable autoregulatory motif.

RESULTS

STILT: A Stochastic Inference Algorithm Using Tree-Structured Time-Lapse Fluorescence Microscopy Data

We introduce STILT for performing parameter inference and model comparison for stochastic chemical reaction networks from fluorescence microscopy movies of proliferating cells (see STAR Methods for details). STILT requires as input quantitative single-cell time series data derived from time-lapse fluorescence microscopy along with the corresponding cellular lineage trees (Figure 1A). STILT iteratively proposes new "particles," i.e., samples, which approximate the probability distribution of both the unknown latent history of the system (including potentially unobserved species) and model parameters given the observed data. Unlike previous approaches, it couples the bootstrap particle filter (Pitt and Shephard, 1999) to a model of cell division to facilitate inference of stochastic gene regulation models, compute evidences, perform model comparison, and infer parameters.

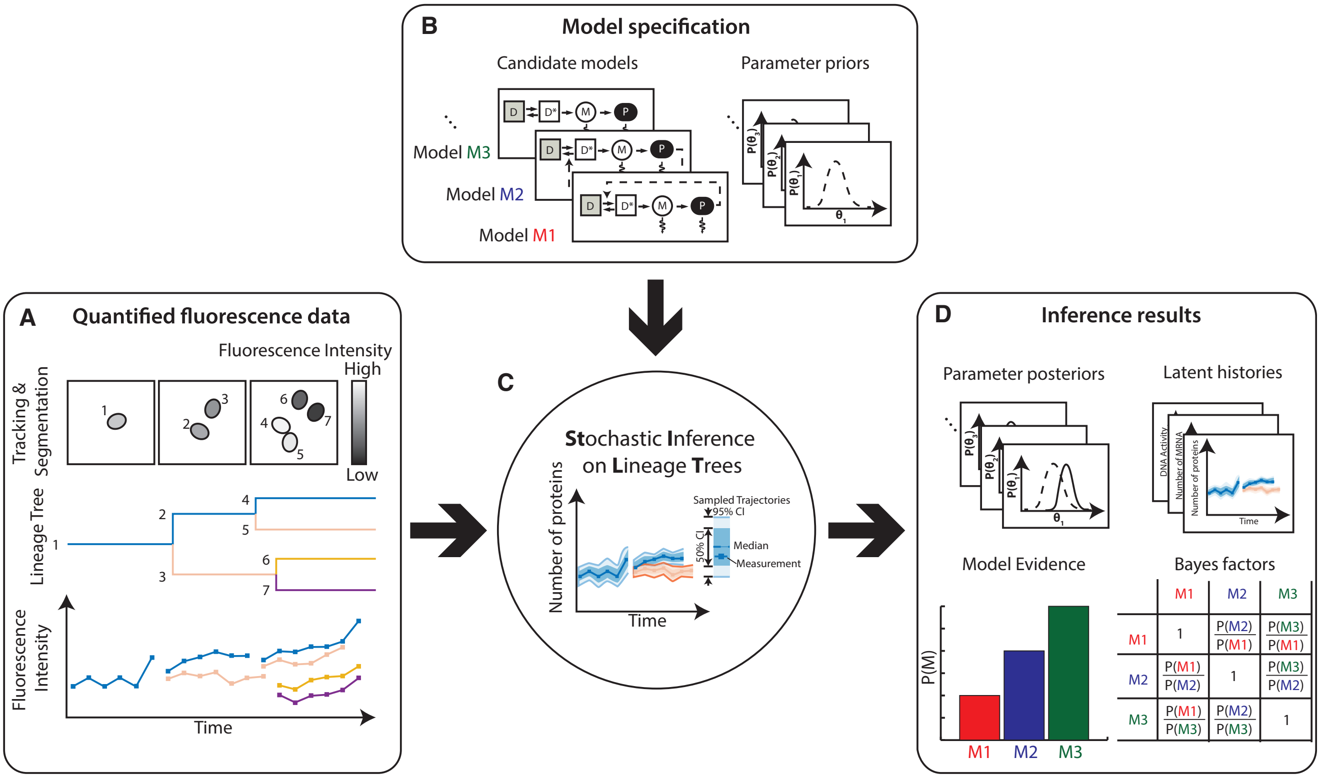
STILT requires the specification of one or more candidate models in the form of chemical reaction networks (Figure 1B), which relate chemical species via their reactions' stoichiometry and chemical kinetic constants. Because the true values of parameters are generally not known, a "prior distribution" (e.g., from literature) is required for each parameter (Figure 1B). STILT (Figure 1C) then combines the experimental data, prior distributions, and candidate mechanistic models to estimate parameters and latent histories and approximate model evidences (Figure 1D). By comparing the evidence for each model, one can compute "Bayes factors" to assess the relative probability of each model and potentially reject unlikely models.

Autoregulatory Motifs for Nanog

We considered three motifs for Nanog autoregulation: no feedback (Figure 2A); negative feedback (Figure 2B); and positive feedback (Figure 2C). In the case of feedback, protein affects the DNA activation/inactivation rate. Each model comprises a single gene in either an inactive conformation (D) with no transcription or an active conformation (D*) with stochastic transcription, mRNA (M), and protein (P). These molecular species are governed by six reactions for the activation/inactivation of DNA and production and degradation of mRNA/protein (Table S1). In the following, we first validated model selection with STILT using synthetic data and subsequently applied STILT to real Nanog time-lapse data.

In Silico Validation

We evaluated STILT on synthetic datasets generated from the above-described autoregulatory motifs (Figures 2A–2C). We simulated each model to yield lineage trees consisting of three generations and seven cells (Figures 2D–2F, solid lines). Parameters were chosen such that cells have similar protein levels (10^3–10^4 molecules) in each model (Table S2). We assume that only protein abundance was measured, thus mRNA and DNA



	M1	M2	M3
M1	1	P(M2)/P(M1)	P(M3)/P(M1)
M2	P(M1)/P(M2)	1	P(M3)/P(M2)
M3	P(M1)/P(M3)	P(M2)/P(M3)	1

Figure 1. Exact Bayesian Inference of Stochastic Gene Regulation Models Using Single-Cell Time-Lapse Fluorescence Lineage Trees

(A) Quantified fluorescence lineage trees are extracted from time-lapse fluorescence microscopy movies.

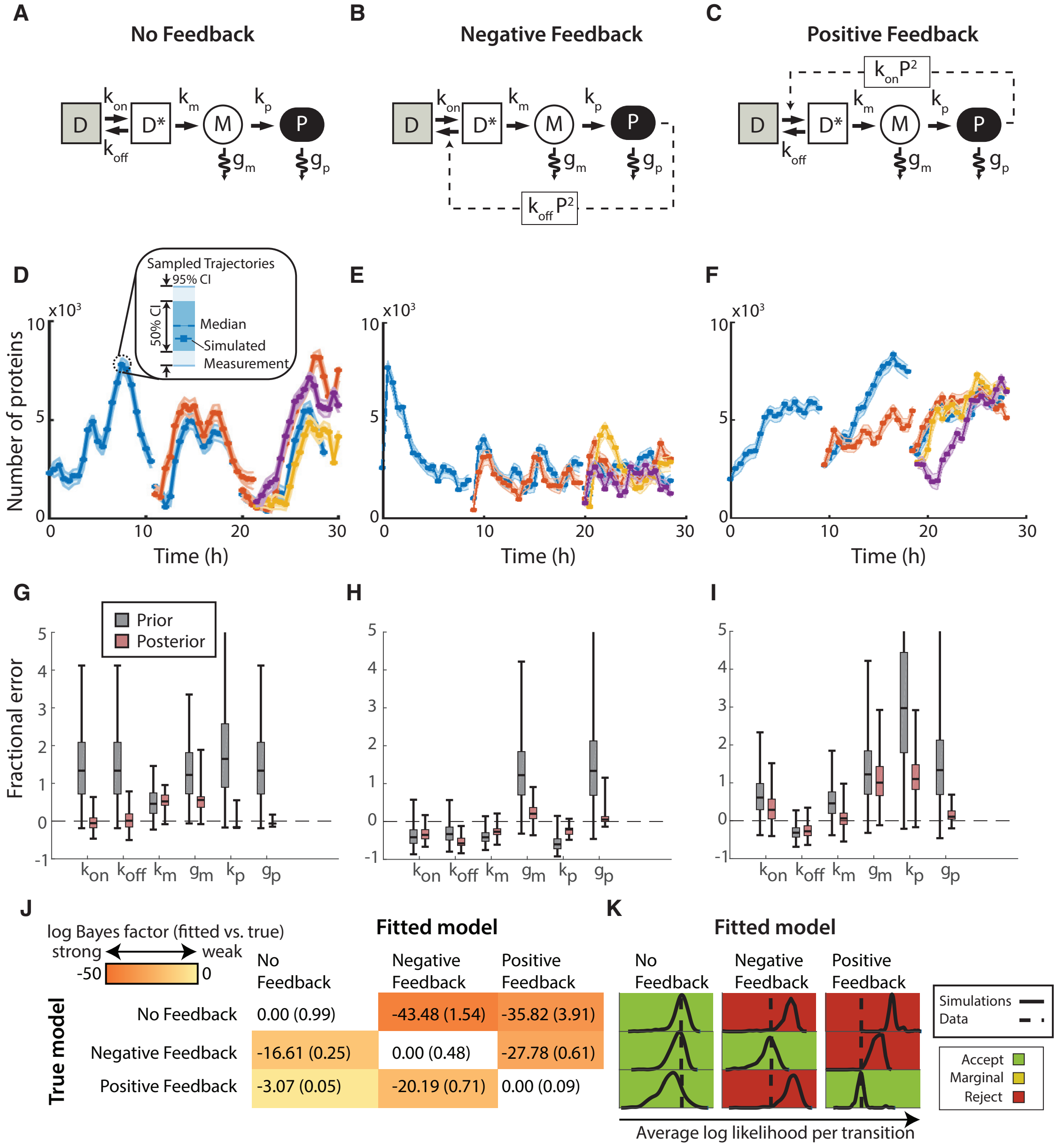
(B and C) The trees are combined with (B) candidate models and their respective parameter prior distributions and serve as input to the particle-filter-based inference algorithm STILT (C).

(D) STILT generates estimates for the posterior distribution of each model parameter, latent histories, and evidence of each model. The latter is used for model comparison using Bayes factor, which is the ratio of the marginal likelihoods of two models.

are unobserved. Gaussian noise ($\sigma = 200$ proteins) was added to simulate measurement error.

We applied STILT to each lineage tree using suitable prior distributions (Figure S5; Table S3), with three runs per dataset to assess robustness. Notably, we find that STILT proposes trajectories that completely contain the observed time series for each simulated dataset (Figures 2D–2F, shaded areas); if sampled trajectories would not contain the observed data, this would strongly suggest that the model and/or prior distributions cannot accurately capture the behavior present in the synthetic datasets. The latent mRNA trajectories are also well inferred by the particle filter (Figure S6). Using fractional errors of each parameter, $(\theta_i - \theta_i^{true})/\theta_i^{true}$, where θ_i^{true} is the true value of parameter *i*, we find general improvement compared to prior distributions (Figures 2G–2I). Whereas many parameters are estimated accurately and well contained in the "posterior distributions," some parameters are poorly identifiable, probably due to insufficient information content of the simulated datasets. For example, parameters k_m and g_m are poorly identified for the no feedback model (Figure 2G). Most parameters are robustly estimated when the model is correct (Figure S1). However, inference with an incorrect model may result in local optima of inferred model parameters, i.e., a non-robust estimate, if the observed transitions are very unlikely to occur for the assumed model.

We evaluated the ability to correctly select the true model from a set of candidate models by computing the model evidence (defined as the marginal log likelihood of the model) of each model/dataset combination. For each dataset, there is one true model that was used to generate the data (e.g., no feedback); the remaining two models (e.g., negative feedback and positive feedback) are, by definition, incorrect. We then computed the log Bayes factors, i.e., differences in the evidence (Figure 2J) for all three models, and found that the known true model is preferred in each case. Typically, a log Bayes factor larger than 3 is considered strong evidence (Kass and Raftery, 1995). For example, the difference in evidence of the (correct) no feedback and (incorrect) negative feedback model (Figure 2J, top row) is -43.48 ± 1.54 (mean ± SD; n = 3 runs), indicating strong preference for the no feedback model over the negative feedback model. Moreover, the log Bayes factor between the true model and the incorrect models is strong (≥ 3) and robust for the true model in each scenario. To illustrate the utility of incorporating cell lineages into our analysis, we compared STILT to a conventional particle-filter-based algorithm that ignores cellular genealogy, instead inferring parameters for each cell independently. In several instances, the true model was not identified, and the Bayes factors were generally smaller when neglecting genealogy (Figure S7; Table S4).



True model	No Feedback	Negative Feedback	Positive Feedback
No Feedback	0.00 (0.99)	-43.48 (1.54)	-35.82 (3.91)
Negative Feedback	-16.61 (0.25)	0.00 (0.48)	-27.78 (0.61)
Positive Feedback	-3.07 (0.05)	-20.19 (0.71)	0.00 (0.09)

Figure 2. STILT Correctly Identifies Autoregulatory Models in Synthetic Data

(A–C) We consider three simple models of transcriptional control: (A) no feedback; (B) negative feedback; and (C) positive feedback. Models differ in the propensity of DNA (D/D*) activation and inactivation. Further components of the system comprise mRNA (M) and protein (P) (see Table S1 for details on system reactions).

(D–F) We simulate each model to generate quantified lineage trees of measured protein numbers, and subsequently perform inference using STILT. The median (dashed line) and 50%, 95% confidence intervals of the trajectories sampled by the particle filter (band plots) show excellent agreement with the simulated data (squares). Results shown for the (D) no feedback, (E) negative feedback and (F) positive feedback models.

(G–I) STILT also estimates posterior distributions of model parameters (red, 99% confidence interval). For many parameters the posterior distribution shows improved estimates compared to the prior distribution (gray, 99% confidence interval) in terms of the fractional error, defined as the error of each parameter sample divided by the true value of that parameter; a fractional error of zero indicates a perfect inference result. Parameter inference results shown for the (G) no feedback, (H) negative feedback and (I) positive feedback models.

(J) Log Bayes factors (mean, SD; n = 3 inference runs), i.e., the difference in the marginal log likelihood of each model from that of the true model for each dataset, indicate that the correct model is always strongly preferred (white; diagonal).

(K) The goodness-of-fit test (see STAR Methods) approximates the distribution of average log likelihood per transition for simulations generated using the inferred parameters for each model (solid line). If the average log likelihood of the actual dataset (dashed line) falls within this distribution, it indicates good agreement of the dataset with that simulated model.

Although Bayes factors facilitate model selection, it is not in general possible to determine whether a preferred model is "compatible" with a particular, experimentally generated dataset, i.e., whether the data could have realistically been generated by that model with the inferred parameters. Thus, we developed a simple test to assess the correctness of the inferred model and parameters. Specifically, we compared the likelihood of observing the real data (computed using the inferred model and parameters) with the likelihood of observing synthetic datasets generated using the same model and parameters. We categorized each model as either reject (test statistic outside 98% confidence interval), marginal (outside 95% confidence interval), or accept (within 95% confidence interval). Our "goodness-of-fit test" accepts the true model for each dataset (Figure 2K; diagonal). By contrast, the goodness-of-fit test rejects the negative feedback model fit to the positive feedback and no feedback datasets and the positive feedback model fit to the negative feedback and no feedback datasets, from which we can deduce that the model is unlikely to be true for that dataset. However, the no feedback model shows agreement with simulated datasets from both the negative and positive feedback models. This level of agreement is likely due to the less constrained expression dynamics of the no feedback model compared to the other models.

Inference of Nanog Autoregulatory Motifs Using STILT Rejects Positive Autoregulation

We next applied STILT to study the debated autoregulation mechanism of Nanog, using time-lapse data from a recent single-cell study (Filipczyk et al., 2015). In these experiments, the fluorescence intensity of NanogVENUS, a reporter for the protein expression of the pluripotency factor Nanog, was quantified for single cells over several generations, thus providing cellular genealogies. To facilitate computation, we then extracted smaller non-overlapping "subtrees" of three generations (seven cells) each for analysis with STILT (Figure 3A). We converted fluorescence intensities in 15 subtrees to absolute protein numbers (see STAR Methods) and performed minimal data cleaning to remove incorrectly segmented or quantified measurements (Figure S8). Finally, we used STILT to perform inference with the three autoregulatory motifs introduced above. Prior distributions for each model parameter were estimated from available knowledge (Table S5).

STILT produced sampled trajectories that agree well with the measured time series (Figures 3B and 3C show one subtree fit with the no feedback model; see Mendeley data archive [STAR Methods] for all models and subtrees), indicating that all models are capable of reproducing the observations with the assumed parameter distributions, albeit with varying likelihoods. The estimated latent mRNA abundances (Figure 3C) agree well with recent estimates of approximately 100–300 copies per cell (Nair et al., 2015; Singer et al., 2014). We found that the subtrees are informative in the sense that they influence the estimated probability distributions for model parameters, i.e., parameter posterior distributions are shifted relative to their prior distributions (Figures 3D and S3; Table S6). Moreover, parameters are robustly estimated (i.e., the posterior distributions are consistent) over three technical replicates (Figure S3). Next, we estimated the evidence of each model for each subtree (Table S7). We found that the no feedback model is preferred in most cases (11/15) and is significantly greater than the next best model in ten of these instances (Figure 3E). For four subtrees, the negative feedback model is preferred, and the evidence is significantly greater in two of these. By contrast, it is consistently much lower for the positive feedback model.

Finally, we used the goodness-of-fit test to assess the ability of each model to explain the data. We found that both the no feedback and negative feedback models agree well with the observed datasets when using the median of the estimated posterior distributions (Figure 3F; Nanog goodness-of-fit plots in Mendeley data archive [STAR Methods]; Table S8). The negative feedback model is compatible with the most subtrees (13/15 accepted); in contrast, only 8/15 subtrees were compatible with the no feedback model (five were marginally accepted). However, the positive feedback model is accepted for only 5/15 subtrees and marginally for one additional subtree. For two subtrees, no model could be rejected, and all models are rejected in terms of the goodness of fits for subtree number 14 (Table S8).

Model-Based Experimental Design for Selection of Nanog Autoregulation Motif

Based on the goodness-of-fit test and Bayes factors analysis, we can reject positive feedback as a putative motif for Nanog autoregulation for the analyzed datasets. Note that this rejection is based on statistical reasoning—positive feedback with the specific structure and parameter priors assumed does not explain the data well. To discriminate between the remaining two alternatives, we used STILT to devise an experiment whose outcome would differ significantly for the no/negative feedback models. We consider the inclusion of exogenous expression of transgenic Nanog, which would increase the effective rate of DNA inactivation in the negative feedback model (Figure 4A). To quantify the predicted effect of this perturbation, we simulated negative feedback using the previously inferred parameters, while introducing varying levels of exogenous Nanog (P_{ex}). We found a strong shift in endogenous Nanog dynamics at only a few hundred thousand molecules of exogenous Nanog and complete downregulation for $P_{ex} > 10^6$ (Figure 4B). Thus, by introducing exogenous Nanog into the experimental system, we would expect decreased levels of endogenous Nanog if the negative feedback model and inferred model parameters accurately capture the observed biological behavior; by contrast, if the no feedback model is correct, we would expect constant levels of endogenous Nanog (evaluated at 46 hr; Figure 4C).

We tested our model prediction using an mESC line with fluorescent reporters for both endogenous and exogenous Nanog (Figure S4A). We quantified exogenous Nanog and defined five compartments of expression: no exogenous; 1×, 2×, and 3× overexpression (OE); and "very high" (i.e., >3× OE), relative to endogenous expression in untransfected cells (Figures S4B and S4C). In agreement with recent reports (Fidalgo et al., 2012; Navarro et al., 2012), expression of transgenic exogenous Nanog was found to induce a dose-dependent downregulation of endogenous Nanog production (Figure 4D). We then replicated the experimental perturbation using STILT. Using the estimated parameters, we simulated exogenous Nanog corresponding to the quantity of exogenous Nanog of each

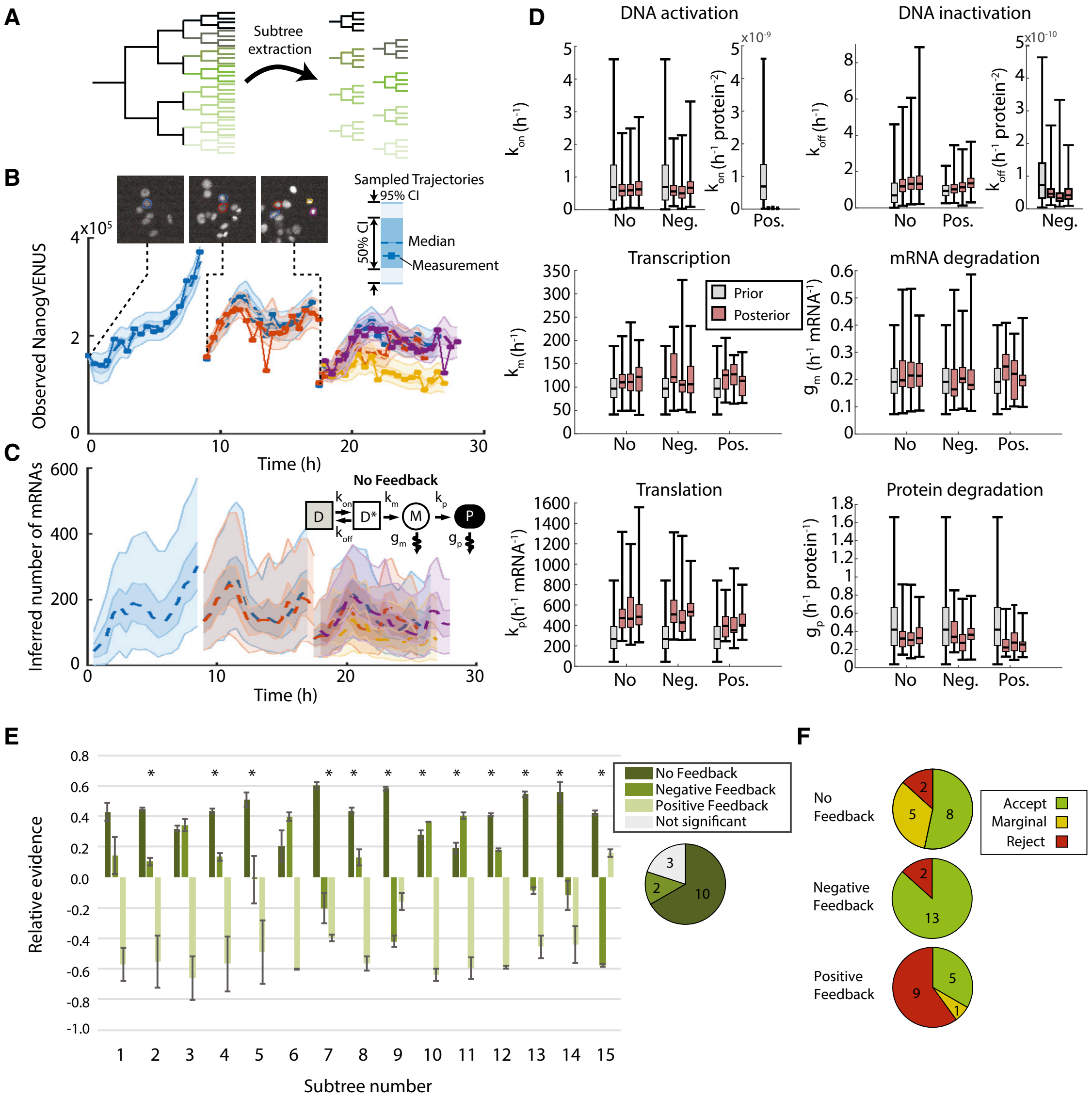


Figure 3. Model Comparison Suggests that NanogVENUS Expression Dynamics in ESCs Are Best Explained by the No Feedback or Negative Feedback Model

(A) Fluorescence lineage trees are partitioned to obtain disjoint subtrees of seven cells each.

(B) STILT yields samples for the latent trajectories of proteins that reproduce and contain the observed NanogVENUS data (shown for the no feedback model). Colony images are shown at the first time point of each generation.

(C) The latent history of mRNA is inferred and agrees with previous estimates of the mRNA copy number of Nanog in ESCs (shown for the no feedback model).

(D) We compare the estimated posterior distributions of model parameters for the same subtree fit with each of the models. Each dataset is fit three times (red; 99% confidence interval), and the resulting posterior distributions compared with the prior distributions (gray; 99% confidence interval). We find that estimates for mRNA and protein parameters are robust between technical replicates and across models.

(E) We compute the evidence of each model, shown relative to the average over all models for that subtree (mean, SEM; n = 3), scaled by evidence range for that subtree (see Table S7 for absolute values). We find that the no feedback model provides the largest evidence in most cases (*, significant with $p < 0.01$), whereas the negative feedback model is preferred for four subtrees. The evidence for the positive feedback model is generally lower than the other models. The frequency for which the evidence for each model is significantly greater than the remaining models is shown with a pie chart.

(F) The goodness-of-fit test indicates that the negative feedback model is accepted for most subtrees (13/15) compared to 8/15 for no feedback and 5/15 for positive feedback. Each model is rated as accept, marginal, or reject based on the result of the goodness-of-fit test.

overexpression compartment. We found agreement between the predicted and measured decrease in endogenous Nanog expression levels upon perturbation (Figure 4E). Note that the prediction uses only parameters inferred from the time-lapse data and the estimated quantities of exogenous Nanog. The agreement suggests that the data are well explained by negative feedback with the assumed mechanistic model and not by the no feedback model.

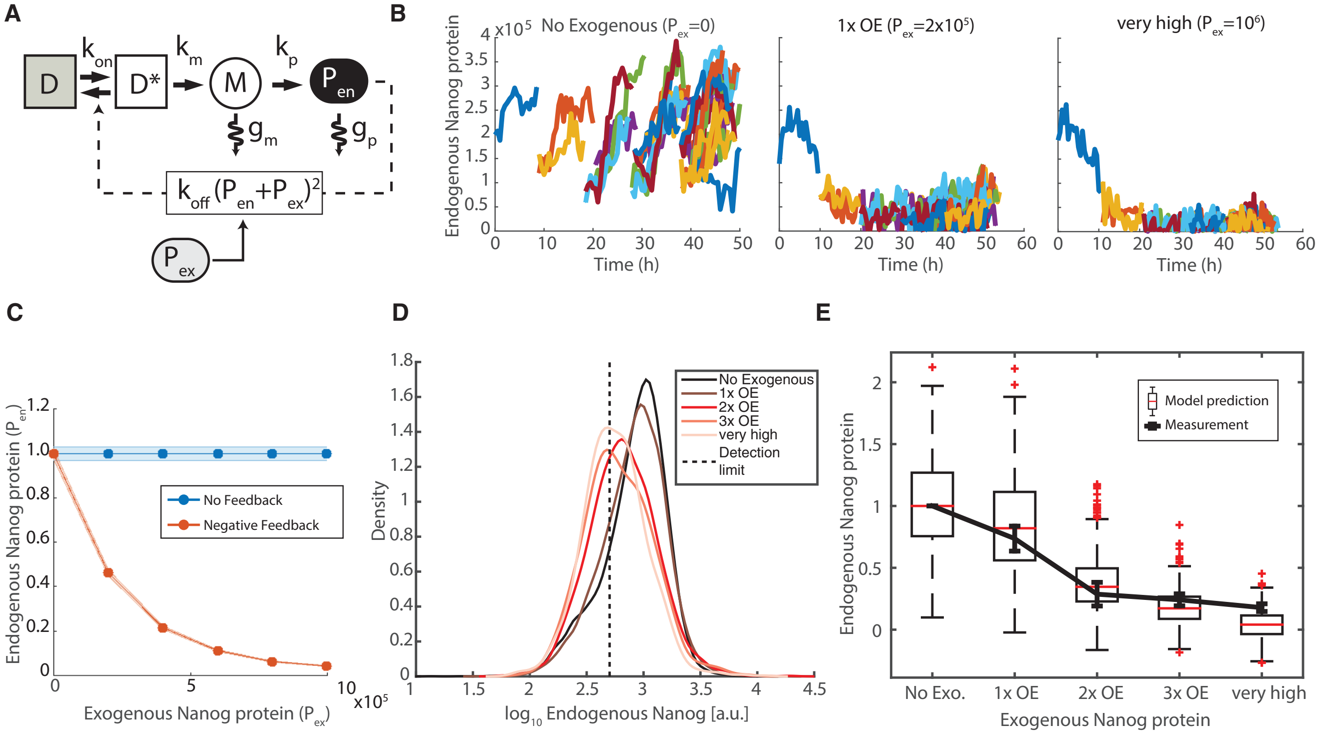


Figure 4. We Experimentally Verify the Predicted Response of the Negative Feedback Model to Nanog Overexpression

(A) We modify the negative feedback model to incorporate exogenous Nanog (P_{ex}), which acts with endogenous Nanog (P_{en}) to increase the propensity of DNA inactivation.

(B) Using the previously inferred model parameters, we generate synthetic trees for various levels of exogenous Nanog, illustrated for 0 molecules, 2 × 10^5 molecules (100% increase), and 10^6 molecules (500% increase).

(C) We predict strong downregulation (fold-change relative to median expression of unperturbed cells) of endogenous Nanog (mean, ±2 SEM; n = 30 simulations) for the negative feedback model; the no feedback model is unperturbed by exogenous Nanog.

(D) Endogenous Nanog levels decrease as the amount of exogenous Nanog increases. Detection threshold is shown as a dashed line (see Figure S4B).

(E) Using the negative feedback model with exogenous Nanog, we compare the predicted fold-change (box and whiskers) in endogenous Nanog in response to exogenous Nanog overexpression with the experimentally determined median fold-change (mean, SEM of three replicates; line). Fold-change is relative to median expression level of endogenous Nanog in the unperturbed (no exogenous) compartment.

Comparative Validation of Estimated Parameters for the Negative Feedback Model

STILT yields estimates of parameters, including the rate of switching between active and inactive DNA conformations, transcription and translation rates, and degradation rates of mRNA and protein. These estimates agree well with previous estimates. The inferred median mRNA degradation rate for each subtree ranges from 0.11 to 0.43 (mean 0.23; n = 15) per molecule per hour, which agrees well with the previous estimate of 0.15 (Ochiai et al., 2014). Estimated protein degradation rates range from 0.03 to 0.44 (mean 0.22) per molecule per hour. This estimate is consistent with the previous estimates of 0.14–0.35 (Figure 3D), which correspond to estimated half-lives of 5 hr and 2 hr, respectively (Abranches et al., 2013; Filipczyk et al., 2015). Thus, both Nanog protein and mRNA have comparable half-lives of ~3 hr. Transcription rates range from 67.2 to 181.8 (mean 110.3) per hour, consistent with the estimate of 126.6 per hour (Ochiai et al., 2014). Translation rates range from 215.9 to 1,142.0 per mRNA per hour. This quantity is not well characterized in literature but agrees roughly with the estimate of up to 1,000 estimated for mouse fibroblasts (Schwanhäusser et al., 2011, 2013). The mean value of these estimates across subtrees is similar between the no feedback and negative feedback models: 115.0 versus 110.3 for translation; 0.26 versus 0.23 for mRNA degradation; 637.8 versus 619.5 for translation; and 0.24 versus 0.22 for protein degradation for the no feedback and negative models, respectively.

DNA activation and inactivation rates cannot be easily assessed because they represent an abstraction of a more complicated biochemical process. For example, activation might correspond to changes in the DNA and histone modification state of the promoter, which permit greater transcriptional activity (Singer et al., 2014). Nonetheless, the estimated rate of activation ranges from 0.27 to 1.74 (mean 0.68) per hour, which is consistent with the estimate of 1.69 per hour in the simple unregulated telegraph model of Ochiai et al. (2014). The inactivation rate ranges from 0.13 × 10^{-11} to 1.12 × 10^{-10} (mean 5.69 × 10^{-11}) per hour. In the negative feedback model, this rate scales quadratically with the number of proteins to give a total rate of approximately 2.5–5.0 per hour (assuming 2–4 × 10^5 Nanog protein molecules per cell). This estimate is substantially smaller than the estimate of 36.54 per hour in the telegraph model (Ochiai et al., 2014). However, there the model assumes DNA to be inactive whenever active transcription is not detected. In contrast, the stochastic nature of our model allows DNA to remain in the active state even between transcription events, which may contribute to a reduced overall rate of DNA inactivation. We also note that the estimated number of mRNAs inferred by STILT, which ranges from approximately 0 to 300, agrees well with previous estimates of ~100 ± 100 (Hansen and van Oudenaarden, 2013; Nair et al., 2015; Singer et al., 2014; see Nanog Bandplots in Mendeley data archive [STAR Methods]). In summary, STILT achieves comprehensive rate constant estimates of the different processes governing Nanog dynamics solely from a time-lapse study. These are in good agreement with the results from various dedicated studies, each independently focusing on selected aspects, such as DNA (in-) activation or mRNA/protein synthesis and degradation.

DISCUSSION

As our analysis of real and simulated data with STILT demonstrates, fitting mechanistic models to time-lapse data facilitates the analysis of latent variables and enables the design of informative experiments. The sampled trajectories may also provide valuable insight into the dynamics of latent variables, including DNA activity and mRNA copy number. The inferred trajectories can also be analyzed to provide information about gene activity, such as inferring continuous versus burst transcription, possible oscillations, refractory periods, etc. (Suter et al., 2011). For example, examining the mRNA trajectories (see Nanog Bandplots in Mendeley data archive), we observe both burst-like and sustained transcriptional modes.

In this work, we use this approach to investigate regulatory motifs. A variety of hypothetical motifs for Nanog regulation have previously been proposed that aim to recapitulate Nanog's prominently heterogeneous expression (Chambers et al., 2007; Filipczyk et al., 2013). Such motifs can produce phenomena, such as bistability or oscillations (Glauche et al., 2010), and excitatory excursions from a stable state (Kalmar et al., 2009), which may give rise to broad or bimodal Nanog distributions resembling those observed in static snapshot experiments. On the other hand, Nanog transcriptional dynamics have been described statistically using a simple, unregulated telegraph model fit to the timing of periods of gene activity (Ochiai et al., 2014). To ultimately discriminate between such mechanisms, quantitative model selection based on single-cell data is required. However, until now, extracting mechanistic knowledge from fluorescent fusion protein trajectories has been hampered by the lack of suitable inference techniques. In particular, the intrinsic stochasticity of Nanog expression at the single cell level and the proliferating nature of mESC populations necessitate an approach that is fully stochastic, Bayesian, and suited to tree-structured data. Using STILT, we overcome these challenges to make use of the full information content of time-lapse fluorescence movies, quantitatively fit and select among putative models of autoregulation.

Consistent with recent evidence, we find positive feedback to be unlikely to explain the observed Nanog lineage trees. Instead, STILT indicates greater evidence for the no feedback model for many subtrees and negative feedback for fewer subtrees, whereas positive feedback is consistently disfavored. However, the goodness-of-fit test indicates overall better agreement with data for the negative feedback model. The stronger evidence for no feedback arises because the fitted parameter values are a priori more likely with the assumed prior distributions compared to the negative feedback model, i.e., the negative feedback model agrees with the data for a more limited set of parameters, which were assumed to be less likely. However, as for all Bayesian inference methods, this result is influenced by the choice of prior distributions and thus should be considered in context of the goodness-of-fit test results.

To discriminate between no feedback and negative feedback, we used STILT as an experimental design tool and quantitatively predicted the strength of downregulation upon overexpression. Further investigation using novel experiments revealed the expected strong downregulation upon high expression of transgenic Nanog, in very good agreement with model predictions. Taken together, we conclude that Nanog negative autoregulation is indeed likely but has a prominent effect only at relatively high levels of protein expression, which renders model discrimination based on Bayes factors alone difficult. The lack of strong autoregulation suggests stable oscillations to be unlikely, in accordance with previous analysis (Filipczyk et al., 2015), and supports the notion that Nanog undergoes broad fluctuations that serve to diversify the mESC population's ability to respond to differentiation cues (Abranches et al., 2014; Torres-Padilla and Chambers, 2014).

The inferred motif naturally represents a simplification of Nanog's true regulatory mechanism. For example, although Nanog autoinhibition is thought to be mediated by Zfp281 and the NuRD complex (Fidalgo et al., 2012), these factors are omitted for simplicity; this is equivalent to assuming Zfp281 abundance to be approximately constant. We further neglect the possibility of monoallelic expression of Nanog. However, it has been previously shown that both Nanog mRNA and protein are highly correlated between alleles (Filipczyk et al., 2013; Hansen and van Oudenaarden, 2013), motivating this assumption. Despite these simplifications, the negative feedback model (1) produces sample trajectories that reproduce the observed data, (2) agrees quantitatively with observed fluorescence lineage trees using the goodness-of-fit test, and (3) accurately predicts the magnitude of downregulation in overexpression experiments. Thus, we conclude that the autoinhibitory motif provides a simple but accurate description of Nanog protein dynamics, superior to the considered alternatives.

Whereas we have focused on Nanog autoregulation, STILT may be used for inference and model selection for arbitrary stochastic gene regulation models applied to fluorescence lineage trees, thus enabling quantitative and exact analysis of lineage-tracked time-lapse fluorescence data. Future research will focus on improving STILT's performance for large/high-dimensional datasets. The generic MATLAB implementation is provided as open source with SBML compatibility for easy import of user-specified models.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

    - ○ Exogenous Nanog Construct
- ● METHOD DETAILS
    - ○ Chemical Reaction Networks
    - ○ Inference of Latent History and Model Parameters
    - ○ Model Assumptions
    - ○ Model Specifications
    - ○ Model Evaluation
    - ○ Time-Lapse Fluorescence Microscopy Data
    - ○ Experimental Validation of Negative Feedback Model
- ● DATA AND SOFTWARE AVAILABILITY
    - ○ Software
    - ○ Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.11.001.

AUTHOR CONTRIBUTIONS

Conceptualization, J.F., C.M., F.J.T., M.C., and T.S.; Supervision, M.C., C.M., and F.J.T.; Methodology, J.F.; Software, S.G. and J.F.; Investigation, J.F. and S.H.; Visualization, J.F.; Formal Analysis, J.F.; Writing – Original Draft, J.F., C.M., and M.C.; Writing – Review & Editing, J.F., C.M., and M.C.; Data Curation, J.F. and M.S.; Resources, M.C., T.S., M.S., A.F., and S.H.; Funding Acquisition, M.C., F.J.T., and T.S.

ACKNOWLEDGMENTS

We thank Virginia-Cezara Luca for single-cell quantification, Will MacNair for feedback on the manuscript and analysis suggestions, and Jan Hasenauer and Michael Strasser for helpful comments and insightful discussions. We also thank the Helmholtz Zentrum München, ETH Zürich, and the ERC (starting grant Latent-Causes to F.J.T. grant number: 259294) and the SystemsX.ch (RTD HDL-X) for generous funding support.

Received: June 1, 2016
Revised: August 25, 2016
Accepted: October 31, 2016
Published: November 23, 2016

REFERENCES

Abranches, E., Bekman, E., and Henrique, D. (2013). Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of Nanog expression. PLoS ONE *8*, e59928.

Abranches, E., Guedes, A.M.V., Moravec, M., Maamar, H., Svoboda, P., Raj, A., and Henrique, D. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. Development *141*, 2770–2779.

Becskei, A., Kaufmann, B.B., and van Oudenaarden, A. (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat. Genet. *37*, 937–944.

Bornstein, B.J., Keating, S.M., Jouraku, A., and Hucka, M. (2008). LibSBML: an API library for SBML. Bioinformatics *24*, 880–881.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Cao, Y., Gillespie, D.T., and Petzold, L.R. (2006). Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys. *124*, 044109.

Cao, Y., Gillespie, D.T., and Petzold, L.R. (2007). Adaptive explicit-implicit tau-leaping method with automatic tau selection. J. Chem. Phys. *126*, 224101.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Coutu, D.L., and Schroeder, T. (2013). Probing cellular processes by long-term live imaginghistoric problems and current solutions. J. Cell Sci. *126*, 3805–3815.

Doucet, A., and Johansen, A.M. (2009). A tutorial on particle filtering and smoothing: fifteen years later. Handbook Nonlinear Filtering *12*, 656–704.

Dunlop, M.J., Cox, R.S., Levine, J.H., Murray, R.M., and Elowitz, M.B. (2008). Regulatory activity revealed by dynamic correlations in gene expression noise. Nat. Genet. *40*, 1493–1498.

Eilken, H.M., Nishikawa, S., and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature *457*, 896–900.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183–1186.

Fidalgo, M., Faiola, F., Pereira, C.-F., Ding, J., Saunders, A., Gingold, J., Schaniel, C., Lemischka, I.R., Silva, J.C.R., and Wang, J. (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. Proc. Natl. Acad. Sci. USA *109*, 16202–16207.

Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P.S., Lickert, H., Anastassiadis, K., and Schroeder, T. (2013). Biallelic expression of nanog protein in mouse embryonic stem cells. Cell Stem Cell *13*, 12–13.

Filipczyk, A., Marr, C., Hastreiter, S., Feigelman, J., Schwarzfischer, M., Hoppe, P.S., Loeffler, D., Kokkaliaris, K.D., Endele, M., Schauberger, B., et al. (2015). Network plasticity of pluripotency transcription factors in embryonic stem cells. Nat. Cell Biol. *17*, 1235–1246.

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

Gillespie, D.T. (2007). Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. *58*, 35–55.

Glauche, I., Herberg, M., and Roeder, I. (2010). Nanog variability and pluripotency regulation of embryonic stem cellsinsights from a mathematical model analysis. PLoS ONE *5*, e11238.

Golightly, A., and Gillespie, C.S. (2013). Simulation of stochastic kinetic models. Methods Mol. Biol. *1021*, 169–187.

Gordon, N.J., Salmond, D.J., and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F Radar Signal Process *140*, 107–113.

Hansen, C.H., and van Oudenaarden, A. (2013). Allele-specific detection of single mRNA molecules in situ. Nat. Methods *10*, 869–871.

Hilsenbeck, O., Schwarzfischer, M., Skylaki, S., Schauberger, B., Hoppe, P.S., Loeffler, D., Kokkaliaris, K.D., Hastreiter, S., Skylaki, E., Filipczyk, A., et al. (2016). Software tools for single-cell tracking and quantification of cellular and molecular properties. Nat. Biotechnol. *34*, 703–706.

Hoppe, P.S., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K.D., Hilsenbeck, O., Moritz, N., Endele, M., Filipczyk, A., Gambardella, A., Ahmed, N., et al. (2016). Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. Nature *535*, 299–302.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics *19*, 524–531.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567–582.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol. *7*, e1000149.

Kass, R.E., and Raftery, A.E. (1995). Bayes factors. J. Am. Stat. Assoc. *90*, 773–795.

Kreutz, C., Bartolome Rodriguez, M.M., Maiwald, T., Seidl, M., Blum, H.E., Mohr, L., and Timmer, J. (2007). An error model for protein quantification. Bioinformatics *23*, 2747–2753.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science *332*, 475–478.

Locke, J.C.W., and Elowitz, M.B. (2009). Using movies to analyse gene circuit dynamics in single cells. Nat. Rev. Microbiol. *7*, 383–392.

MacArthur, B.D., Sevilla, A., Lenz, M., Müller, F.-J., Schuldt, B.M., Schuppert, A.A., Ridden, S.J., Stumpf, P.S., Fidalgo, M., Maayan, A., et al. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. Nat. Cell Biol. *14*, 1139–1147.

Mullin, N.P., Yates, A., Rowe, A.J., Nijmeijer, B., Colby, D., Barlow, P.N., Walkinshaw, M.D., and Chambers, I. (2008). The pluripotency rheostat Nanog functions as a dimer. Biochem. J. *411*, 227–231.

Munsky, B., and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. J. Chem. Phys. *124*, 044104.

Muzzey, D., and van Oudenaarden, A. (2009). Quantitative time-lapse fluorescence microscopy in single cells. Annu. Rev. Cell Dev. Biol. *25*, 301–327.

Nair, G., Abranches, E., Guedes, A.M.V., Henrique, D., and Raj, A. (2015). Heterogeneous lineage marker expression in naive embryonic stem cells is mostly due to spontaneous differentiation. Sci. Rep. *5*, 13339.

Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N.P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M., and Chambers, I. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. EMBO J. *31*, 4547–4562.

Ochiai, H., Sugawara, T., Sakuma, T., and Yamamoto, T. (2014). Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci. Rep. *4*, 7125.

Pitt, M.K., and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. J. Am. Stat. Assoc. *94*, 590–599.

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

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

Rieger, M.A., Hoppe, P.S., Smejkal, B.M., Eitelhuber, A.C., and Schroeder, T. (2009). Hematopoietic cytokines can instruct lineage choice. Science *325*, 217–218.

Saunders, A., Faiola, F., and Wang, J. (2013). Concise review: pursuing self-renewal and pluripotency with the stem cell factor Nanog. Stem Cells *31*, 1227–1236.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2013). Corrigendum: global quantification of mammalian gene expression control. Nature *495*, 126–127.

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S.H. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. *16*, 45–58.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol. Cell *55*, 319–331.

Sokolik, C., Liu, Y., Bauer, D., McPherson, J., Broeker, M., Heimberg, G., Qi, L.S., Sivak, D.A., and Thomson, M. (2015). Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. Cell Syst. *1*, 117–129.

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

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472–474.

Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. Nature *466*, 267–271.

Torres-Padilla, M.-E., and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development *141*, 2173–2181.

Wang, J., Levasseur, D.N., and Orkin, S.H. (2008). Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. USA *105*, 6326–6331.

Wilkinson, D.J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. Nat. Rev. Genet. *10*, 122–133.

Wilkinson, D.J. (2011). Stochastic Modelling for Systems Biology, Second Edition (CRC Press).

Zechner, C., Pelet, S., Peter, M., and Koeppl, H. (2011). Recursive Bayesian estimation of stochastic rate constants from heterogeneous cell populations. 2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC) 5837–5843.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Lipofectamine 2000	Life Technologies	11668-019
Experimental Models: Cell Lines		
Mouse embryonic stem cell: R1 NanogVENUS	Filipczyk et al., 2015	N/A
Mouse embryonic stem cell: R1 NanogKATUSHKA	Filipczyk et al., 2013	N/A
Recombinant DNA		
Plasmid PB.CAG.iRFPnucmemP2ANanogVENUS	This paper	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

As Lead Contact, Manfred Claassen is responsible for all reagent and resource requests. Please contact Manfred Claassen at claassen@imsb.biol.ethz.ch. with requests and inquiries.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Exogenous Nanog Construct

ESC Culture

Mouse ESCs were of R1 background and report endogenous Nanog protein levels from one allele by fluorescent fusion proteins, either NanogVENUS (NV) or NanogKATUSHKA (NK). ESCs were cultured in DMEM (Catalogue number: D1145 Sigma, MO, USA) supplemented with 2mM GlutaMAX (Catalogue number: 35050-038, GIBCO, USA), 1% Non-essential amino acids (Catalogue number: 11140-035, GIBCO, CA, USA), 1mM Sodium Pyruvate (Catalogue number: S8636, Sigma, MO, USA), 50uM b-mercaptoethanol (Catalogue number: M6250, Sigma-Aldrich, USA), 10% FCS (Catalogue number: 2602P250915, PAN, Aidenbach, Germany) and 10ng/ml LIF (GFM200, Cell Guidance Systems, Cambridge, UK) on 0.1% porcine gelatin (Sigma, Catalogue number: G1890-100G).

Nanog Overexpression Experiments

30,000 cells were seeded in a single well of a 24w plate and transfection was performed 5-7h after seeding. For transfections, 250ng plasmid, 1ul Lipofectamine 2000 (11668-019, Life Technologies) and 50ul Opti-MEM (31985-062; Life Technologies) were mixed and added to the cells. Cells were analyzed by flow cytometry 46h after transfection using a BD LSR Fortessa (BD Biosciences, CA, USA) and data were analyzed with FlowJo (OR, USA). Cells were gated for non-debris and singlets using FCS-A, SSC-A and FCS-W channels. Fluorescence channels were compensated using controls that only expressed one of each fluorescent protein. R1 wild-type cells were used as control for cellular autofluorescence. All experiments have been performed in triplicate.

Expression Plasmids

The Nanog coding sequence was cloned in several variants as a 2A construct into a piggybac vector that has been modified to express a fluorescent nuclear membrane tag (iRFPnucmem) from the CAG promoter using the In-Fusion system (Catalogue number: 638911, Takara, Japan). The resulting constructs are supposed to express iRFPnucmem and Nanog proteins in equal abundances. The plasmid CAG.iRFPnucmem-P2A-NanogVENUS was used in overexpression experiments to allow for comparison of exogenous NanogVENUS levels with endogenous NanogVENUS levels of the R1 NanogVENUS cell line. The NanogVENUS plasmid performed identically to positive control plasmids (CAG.iRFPnucmem-P2A-Nanog; with or without ATG for Nanog). An empty vector control (CAG.iRFPnucmem-P2A) was also used during experiments.

METHOD DETAILS

Chemical Reaction Networks

We consider the case of parameter inference and model comparison for stochastic models of gene regulation described by chemical reaction networks. A chemical reaction network consists of a set of chemical species (e.g., DNA, mRNA, protein, etc.) which may interact via a set of chemical reactions such as synthesis, destruction, or modification.

Each reaction is defined by its stoichiometry, i.e., the quantity of each educt consumed and product produced by the reaction, and reaction rate. Reactions are presumed to take place stochastically as a function of the state of the system, i.e., the number of molecules of each species at a given point in time. The probability of a reaction occurring in infinitesimal time, called the reaction propensity, depends on the number of molecules of each educt available, the reaction volume of the cellular compartment where the reaction takes place, and a reaction constant. Zeroth order reactions involve the production of a species with no dependence on an educt, for example due to constitutive production; their reaction propensities are constant. First order reactions proceed with propensity proportional to the number of available molecules of a single educt. Second order reactions involve two species, and their propensity is proportional to the abundances of both involved educts, and so on. The reaction constants k depend on the chemical species involved, temperature and reaction system volume. Reaction propensities are summarized in the table below.

Reaction order	Reaction	Propensity function (a)
Zeroth	$\varnothing \rightarrow X$	k
First	$X \rightarrow Y$	kx
Second	$X+Y \rightarrow Z$	kxy
Second (same species)	$2X \rightarrow Y$	$kx(x-1)/2$
Third	$X+Y+Z \rightarrow A$	$kxyz$

Inference of Latent History and Model Parameters

Inference Using Bootstrap Particle Filtering

STILT builds upon the recursive, simulation-based particle filter first introduced by Pitt and Sheppard (Pitt and Shephard, 1999). The particle filter approximates the posterior distribution of the latent history of all chemical species for each cell, and all model parameters, iteratively updating the approximation by including new observations.

Consider a chemical reaction network with N_s chemical species, of which $N_o \leq N_s$ are observed via measurement. The vector of reaction constants governing all reactions of this network is denoted by $\boldsymbol{\theta}$. We denote by $\boldsymbol{Y}_t$ the measurement at time t, and by $\mathbf{Y}_{0:N} = (\mathbf{Y}_{t_0}, \ldots, \mathbf{Y}_{t_N})$ the set of observations obtained at a series of N discrete measurement time points $t_0, \ldots, t_N$. Each observation $\mathbf{Y}_t \in \mathrm{R}^{N_o}$ consists of the measurements of N_o observed chemical species. We assume that the observations $\mathbf{Y}_t$ constitute noisy measurements of the true unknown state of the system at time t, denoted by $\mathbf{X}_t$. We denote by $\mathbf{X}_{[t_1,t_2]}$ the path (trajectory) of the random variable $\mathbf{X}_t$ from time t_1 to time t_2, and denote by $\mathbf{X}_{0:N} = (\mathbf{X}_{t_0}, \ldots, \mathbf{X}_{t_N})$ the value of $\mathbf{X}_t$ at the measurement time points $t_0, \ldots, t_N$.

The objective of the bootstrap particle filter is to sample from the posterior joint density $P(\mathbf{X}_{[t_0,t_N]}, \boldsymbol{\theta} \mid \mathbf{Y}_{0:N}, M)$ of latent trajectories $\mathbf{X}_{[t_0,t_N]}$ and parameters $\boldsymbol{\theta}$ for a model with index M, given the observed data $\mathbf{Y}_{0:N}$. We drop the model index M for simplicity; when comparing models we will again introduce this notation. The latent trajectories $\mathbf{X}_{[t_0,t_N]}$ are realizations of a stochastic process, and the data $\mathbf{Y}_{0:N}$ represent noisy observations of (a function of) the latent process obtained at discrete times. The posterior joint density depends on the likelihood $P(\mathbf{Y}_{0:N} \mid \mathbf{X}_{[t_0,t_N]}, \boldsymbol{\theta})$ and parameter prior probability distribution $\pi(\boldsymbol{\theta})$ according to Bayes' Law:

$$\begin{aligned} P\left(\mathbf{X}_{[t_0,t_N]}, \boldsymbol{\theta} \mid \mathbf{Y}_{0:N}\right) &= \frac{P\left(\mathbf{Y}_{0:N} \mid \mathbf{X}_{[t_0,t_N]}, \boldsymbol{\theta}\right) P\left(\mathbf{X}_{[t_0,t_N]}, \boldsymbol{\theta}\right)}{P\left(\mathbf{Y}_{0:N}\right)} \\ &= \frac{P\left(\mathbf{Y}_{0:N} \mid \mathbf{X}_{[t_0,t_N]}\right) P\left(\mathbf{X}_{[t_0,t_N]} \mid \boldsymbol{\theta}\right) \pi(\boldsymbol{\theta})}{P\left(\mathbf{Y}_{0:N}\right)} \\ &= \frac{P\left(\mathbf{Y}_{0:N} \mid \mathbf{X}_{0:N}\right) P\left(\mathbf{X}_{[t_0,t_N]} \mid \boldsymbol{\theta}\right) \pi(\boldsymbol{\theta})}{P\left(\mathbf{Y}_{0:N}\right)}. \end{aligned} \tag{1}$$

The simplification on the right side of (Equation 1) is possible since the probability of observing data $\mathbf{Y}_{0:N}$ given a latent trajectory $\mathbf{X}_{[t_0,t_N]}$ depends only on its value at the measurement time points $\mathbf{X}_{0:N}$. Furthermore, it does not depend on the underlying parameters $\boldsymbol{\theta}$ of the stochastic process (measurement error is considered separately). The probability $P(\mathbf{X}_{[t_0,t_N]} \mid \boldsymbol{\theta})$ captures the evolution of the stochastic process parameterized by $\boldsymbol{\theta}$.

The observations $\mathbf{Y}_{0:N}$ are related to $\mathbf{X}_{[t_0,t_N]}$ by a measurement function g with parameter $\boldsymbol{\eta}$: $P(\mathbf{Y}_{0:N} \mid \mathbf{X}_{0:N}) = g(\mathbf{Y}_{0:N}; \mathbf{X}_{0:N}, \boldsymbol{\eta})$, which captures the measurement process and/or apparatus. For example, g might be a (multivariate) Gaussian in which case $\boldsymbol{\eta}$ contains the variance of the measurement process and potentially a scaling factor. We restrict ourselves to chemical reaction networks for which the stochastic process $\mathbf{X}_t$ is a Markov jump process on a subset of the integer lattice N^{N_s}, corresponding to molecular copy numbers reachable by the chemical reactions of the network. For such a system, the exact likelihood of the latent trajectory $P(\mathbf{X}_{[t_0,t_N]} \mid \mathbf{X}_{t_0}, \boldsymbol{\theta})$ can be computed (Wilkinson, 2011), and exact samples of $\mathbf{X}_{[t_0,t_N]} \mid \mathbf{X}_{t_0}, \boldsymbol{\theta}$ can be generated e.g., using Gillespie's algorithm (Gillespie, 1977). Note that the transition density (i.e., $P(\mathbf{X}_{t'} \mid \mathbf{X}_t, \boldsymbol{\theta}), t' > t$) of the stochastic process is in general not known, but can be approximated for small systems, e.g., using the Finite State Projection (Munsky and Khammash, 2006).

Assuming uncorrelated measurement errors in the observation function g, the likelihood $P(\mathbf{Y}_{0:N} \mid \mathbf{X}_{0:N})$ factorizes as:

$$P(\mathbf{Y}_{0:N} \mid \mathbf{X}_{0:N}) = \prod_{i=0}^{N} P\left(\mathbf{Y}_{t_i} \mid \mathbf{X}_{t_i}\right). \tag{2}$$

for a series of N observations.

Furthermore, the stochastic process $\mathbf{X}_t$ is Markovian such that the probability $P(\mathbf{X}_s \mid \mathbf{X}_{[t_0,t]}) = P(\mathbf{X}_s \mid \mathbf{X}_t)$ for some $s \geq t$. The likelihood of the trajectory $\mathbf{X}_{[t_0,t_N]}$ therefore decomposes as:

$$P\left(\mathbf{X}_{[t_0,t_N]} \mid \boldsymbol{\theta}\right) = P\left(\mathbf{X}_{t_0} \mid \boldsymbol{\theta}\right) \prod_{i=1}^{N} P\left(\mathbf{X}_{[t_{i-1},t_i]} \mid \mathbf{X}_{t_{i-1}}, \boldsymbol{\theta}\right) \tag{3}$$

Following the derivation of Gordon et al. (Gordon et al., 1993), we combine Equation 2 and Equation 3, and substitute into Equation 1, to obtain a new expression for the posterior density:

$$
\begin{aligned}
P\left(\mathbf{X}_{[t_0,t_N]},\boldsymbol{\theta} \mid \mathbf{Y}_{0:N}\right) &= \prod_{i=1}^{N} \frac{P\left(\mathbf{Y}_{t_i} \mid \mathbf{X}_{t_i}\right) P\left(\mathbf{X}_{[t_{i-1},t_i]} \mid \mathbf{X}_{t_{i-1}},\boldsymbol{\theta}\right)}{P\left(\mathbf{Y}_{t_i}\right)} P\left(\mathbf{X}_{t_0} \mid \boldsymbol{\theta}\right)\pi(\boldsymbol{\theta}) \\
&= \frac{P\left(\mathbf{Y}_{t_N} \mid \mathbf{X}_{t_N}\right) P\left(\mathbf{X}_{[t_{N-1},t_N]} \mid \mathbf{X}_{t_{N-1}},\boldsymbol{\theta}\right)}{P\left(\mathbf{Y}_{t_N}\right)} \times \prod_{i=1}^{N-1} \frac{P\left(\mathbf{Y}_{t_i} \mid \mathbf{X}_{t_i}\right) P\left(\mathbf{X}_{[t_{i-1},t_i]} \mid \mathbf{X}_{t_{i-1}},\boldsymbol{\theta}\right)}{P\left(\mathbf{Y}_{t_i}\right)} P\left(\mathbf{X}_{t_0} \mid \boldsymbol{\theta}\right)\pi(\boldsymbol{\theta}) \\
&= \frac{P\left(\mathbf{Y}_{t_N} \mid \mathbf{X}_{t_N}\right) P\left(\mathbf{X}_{[t_{N-1},t_N]} \mid \mathbf{X}_{t_{N-1}},\boldsymbol{\theta}\right)}{P\left(\mathbf{Y}_{t_N}\right)} P\left(\mathbf{X}_{[t_0,t_{N-1}]},\boldsymbol{\theta} \mid \mathbf{Y}_{0:N-1}\right)
\end{aligned}
\tag{4}
$$

This can be rewritten as

$$
P\left(\mathbf{X}_{[t_0,t_N]},\boldsymbol{\theta} \mid \mathbf{Y}_{0:N}\right) = w_N \frac{P\left(\mathbf{X}_{[t_{N-1},t_N]} \mid \mathbf{X}_{t_{N-1}},\boldsymbol{\theta}\right)}{P\left(\mathbf{Y}_{t_N}\right)} P\left(\mathbf{X}_{[t_0,t_{N-1}]},\boldsymbol{\theta} \mid \mathbf{Y}_{0:N-1}\right)
\tag{5}
$$

where $w_N = P(\mathbf{Y}_{t_N} \mid \mathbf{X}_{t_N})$.

Hence, there is a simple update rule relating the posterior distribution using observations until time point t_{N-1} to the posterior distribution with the next observation at time point t_N. We note also that one can generate a sample from the posterior joint density of the system at times until t_i, $P(\mathbf{X}_{[t_0,t_i]},\boldsymbol{\theta} \mid \mathbf{Y}_{0:i})$, by first sampling a trajectory from the marginal distibution $P(\mathbf{X}_{[t_0,t_i]} \mid \mathbf{Y}_{0:i})$ and then sampling a parameter $\boldsymbol{\theta} \mid \mathbf{X}_{[t_0,t_i]}$, suggesting a Gibbs sampling approach, i.e., $P(\boldsymbol{X}_{[t_0,t_i]},\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:i}) = P(\boldsymbol{\theta} \mid \boldsymbol{X}_{[t_0,t_i]})P(\boldsymbol{X}_{[t_0,t_i]} \mid \boldsymbol{Y}_{0:i})$.

These observations and the recursive factorization of the joint posterior (Equation 5) motivates the so-called bootstrap (recursive) particle filter (Gordon et al., 1993), which iteratively generates samples (particles) from the posterior distribution conditioned on all prior observations, shown in Algorithm 1.

Algorithm 1: Bootstrap particle filter

```
Data: A set of observed data points $\mathbf{Y}_{0:N}$ at timepoints $t_0,\ldots,t_N$, parameter prior $\pi:\mathbb{R}^d \to \mathbb{R}_{0,+}$,
      observation function $g(\mathbf{Y};\mathbf{X},\boldsymbol{\eta}) = P(\mathbf{Y}|\mathbf{X})$, number of particles $K$
Result: A set of particles $\left\{\left(\mathbf{X}^{(k)}_{[t_0,t_N]},\boldsymbol{\theta}^{(k)}\right)\right\}_{k=1}^{K}$ sampled from the posterior density
      $P(\mathbf{X}_{[t_0,t_N]},\boldsymbol{\theta}|\mathbf{Y})$
Initialization;
for $k$=1...$K$ do
    Sample parameter values from the prior: $\boldsymbol{\theta}^{(k)} \sim \pi(\boldsymbol{\theta})$;
    Sample initial state conditional on first observed data point: $\mathbf{X}^{(k)}_{t_0} \sim \pi(\mathbf{X}_{t_0}|\mathbf{Y}_{t_0},\boldsymbol{\eta})$;
    Initialize particle weight to $w^{(k)}_0 := 1/K$
Iterate over observation time points ;
for $i$=1...$N$ do
    Generate a set of particle indices $\epsilon^{(k)} \in \{1,\ldots,K\}, k=1,\ldots,K$ such that
    $P(\epsilon^{(k)}=a) = w^{(a)}_{i-1}/\sum_{\ell=1}^{K} w^{(\ell)}_{i-1}$;
    for $k = 1\ldots K$ do
        Generate a sample trajectory $\mathbf{X}^{(k)}_{[t_{i-1},t_i]} \sim P(\cdot|\mathbf{X}^{(\epsilon^{(k)})}_{t_{i-1}},\boldsymbol{\theta}^{(\epsilon^{(k)})})$ ;
        Concatenate to previously sampled trajectory: $\mathbf{X}^{(k)}_{[t_0,t_i]} := [\mathbf{X}^{(\epsilon^{(k)})}_{[t_0,t_{i-1}]},\mathbf{X}^{(k)}_{[t_{i-1},t_i]}]$;
        Set the weight of the $k^{\text{th}}$ particle to the likelihood: $w^{(k)}_i := P(\mathbf{Y}_{t_i}|\mathbf{X}^{(k)}_{t_i}) = g(\mathbf{Y}_{t_i};\mathbf{X}^{(k)}_{t_i},\boldsymbol{\eta})$;
        Generate a new set of parameters $\boldsymbol{\theta}^{(k)}$ from the conditional density: $\boldsymbol{\theta}^{(k)} \sim P(\boldsymbol{\theta}|\mathbf{X}^{(k)}_{[t_0,t_i]})$
Sample from the posterior ;
Generate a set of particle indices $\epsilon^{(k)} \in \{1,\ldots,K\}, k=1,\ldots,K$ such that
$P(\epsilon^{(k)}=a) = w^{(a)}_N/\sum_{\ell=1}^{K} w^{(\ell)}_N$;
Construct a sample of $K$ particles from the posterior: $\left\{\left(\mathbf{X}^{(\epsilon^{(k)})}_{[t_0,t_N]},\boldsymbol{\theta}^{(\epsilon^{(k)})}|\mathbf{Y}_{0:N}\right)\right\}_{k=1}^{K}$
```

The recursive particle filter begins by sampling parameters $\boldsymbol{\theta}$ from the parameter prior distribution $\pi(\boldsymbol{\theta})$, and an initial state $\mathbf{X}_{t_0}$ from its prior, for an ensemble of K particles, i.e., each particle is a sample from the joint density of $\mathbf{X}_{t_0}$ and $\boldsymbol{\theta}$. At each iteration i, the particles are resampled according to their normalized weights $w_i^{(k)}/\sum_{\ell=1}^{K} w_i^{(\ell)}$, such that particles that have a state $\mathbf{X}_{t_i}^{(k)}$ for which the current observation $\mathbf{Y}_{t_i}$ is likely are sampled more frequently. After updating the latent history samples $\mathbf{X}_{[t_0,t_i]}^{(k)}$, new samples are generated for the parameters conditional on those latent histories using a Gibbs sampling approach. Together the sample $(\mathbf{X}_{t_i}^{(k)}, \boldsymbol{\theta}^{(k)})$ is used to simulate a new trajectory on the interval $[t_i, t_{i+1}]$ using the stochastic simulation algorithm or variants (Gillespie, 2007; Golightly and Gillespie, 2013). The result of the recursive particle filter at iteration i is an exact sample from the posterior joint density of $(\mathbf{X}_{[t_0,t_i]}, \boldsymbol{\theta} \,|\, \mathbf{Y}_{0:i})$, as shown in Equation 4.

Gamma priors

We consider the case of mass action kinetics models of chemical reaction networks, in which case each parameter $\theta_1, \ldots, \theta_d$ corresponds to the kinetic constant of a chemical reaction (see table above). The inference procedure is significantly simplified if one assumes that the prior distribution of each parameter is gamma distributed, and that the prior distributions of all parameters are conditionally independent:

$$\pi(\boldsymbol{\theta}) = \prod_{p=1}^{d} \pi_p(\theta_p) = \prod_{p=1}^{d} \Gamma(\theta_p; \alpha_p, \beta_p) \tag{6}$$

where α_p and β_p are the hyperparameters for the distribution of $\theta_p, p = 1 \ldots d$, and $\Gamma(x; \alpha, \beta) = (\beta^{\alpha}/\Gamma(\alpha)) x^{\alpha-1} e^{-\beta x}$. The two parameters α and β of the gamma distribution can be chosen e.g., to match a target mean μ and variance σ^2:

$$\alpha \overset{!}{=} \mu^2/\sigma^2 \tag{7}$$

$$\beta \overset{!}{=} \mu/\sigma^2. \tag{8}$$

The assumption of conditional independence of the prior distributions of model parameters is often justified, as information about the covariance of biological constants is often not available.

Using gamma priors for each model parameter, the likelihood $P(\mathbf{X}_{[t,t+\tau]} \,|\, \boldsymbol{\theta})$ of a particular (fully-observed) realization of the Markov jump process $\mathbf{X}_t$ on the interval $[t, t+\tau]$ is conjugate to the prior, such that the posterior probability $P(\boldsymbol{\theta} \,|\, \mathbf{X}_{[t,t+\tau]})$ is also gamma distributed (see Wilkinson et al. (Wilkinson, 2011), p. 281):

$$\begin{aligned} P(\boldsymbol{\theta} \,|\, \mathbf{X}_{[t,t+\tau]}) &= \frac{P(\mathbf{X}_{[t,t+\tau]} \,|\, \boldsymbol{\theta})}{P(\mathbf{X}_{[t,t+\tau]})} \pi(\boldsymbol{\theta}) = \frac{P(\mathbf{X}_{[t,t+\tau]} \,|\, \boldsymbol{\theta})}{P(\mathbf{X}_{[t,t+\tau]})} \prod_{p=1}^{d} \Gamma(\theta_p; \alpha_p, \beta_p) \\ &= \prod_{p=1}^{d} \Gamma(\theta_p; \alpha_p + r_p, \beta_p + G_p) \end{aligned} \tag{9}$$

where r_p is the number of reaction firings of reaction p on the interval $[t, t+\tau]$ and the term $G_p = (1/\theta_p) \int_t^{t+\tau} a_p(\mathbf{X}_s) ds$ is the integral of the reaction propensity a_p of the p^{th} reaction (see table above), rescaled by the reaction constant θ_p. The rescaling renders G_p dependent only on the instantaneous configuration of the system at all points along the trajectory, and not on the reaction constants. Hence, a new sample for $\boldsymbol{\theta}$ given the newly simulated trajectory (line 13 of Algorithm 1) can be generated by simply sampling from the updated gamma posterior (Equation 9); furthermore, the summary statistics r_p and G_p are sufficient for describing the posterior distribution of $\boldsymbol{\theta}_p$. Thus, the full trajectories do not need to be stored, but instead the new simulations can be used to merely update the parameters of the posterior distribution of $\boldsymbol{\theta}$ (i.e., set $\alpha_p \leftarrow \alpha_p + r_p, \beta_p \leftarrow \beta_p + G_p$), reducing storage requirements.

STILT: Stochastic Inference on Lineage Trees

The particle filtering strategy described in Algorithm 1 is suitable for inference of the latent history of a single cell. However, if the cellular lineage is known it is possible to exploit the tree structure to improve the performance of the inference algorithm, for instance by constraining the range of possible initial values for daughter cells at the moment of division according to the state of the mother cell. This is more informative than assuming arbitrary distributions for the initial conditions of latent species, as required in Algorithm 1. Moreover, when incorporating the tree structure, the inferred parameter values are required to generate trajectories that have a high likelihood for multiple cells simultaneously as the cells proliferate.

STILT performs tree-based inference as outlined in Algorithm 2 below. We consider a tree comprised of N_c total cells, with indices $j = 1, \ldots, N_c$. For the j^{th} cell there is a series of N_j (possibly multivariate) measurements obtained at times ${}^{j}t_1, \ldots, {}^{j}t_{N_j}$, corresponding to each of the N_o observed species, denoted by ${}^{j}\mathbf{Y}_{1:N_j} = ({}^{j}\mathbf{Y}_{{}^{j}t_1}, \ldots, {}^{j}\mathbf{Y}_{{}^{j}t_{N_j}})$,. The algorithm begins by initializing a set of K particles for the tree's founding cell with index $j = 1$, where each particle k comprises both a latent initial state ${}^{1}\mathbf{X}_{t_0}^{(k)}$ and a set of parameters $\boldsymbol{\theta}^{(k)}$. All particles are initially equally weighted as $1/K$. The algorithm then iterates through all of the measurement time points, where for simplicity we assume that measurements of all cells are obtained at a regular interval Δt (i.e., ${}^{j}t_{\ell} = n\Delta t \;\; \forall \ell, n \in \mathrm{N}$); however, the method is equally valid for irregular measurement time intervals. At each iteration i, particles are resampled with frequencies proportional to their weights, and cells that are alive/measured at the current time point $i\Delta t$ are simulated one time step using the generative stochastic process with the sampled parameters $\boldsymbol{\theta}^{(k)}$ to generate a sample ${}^{j}\mathbf{X}_{[i\Delta t,(i+1)\Delta t]}^{(k)}$ of the latent history of cell j over the time interval $[i\Delta t, (i+1)\Delta t]$.

If a cell is observed to divide between this time point and the next, the cell's contents are allocated to the daughter cells. If both daughter cells are present, the total cellular contents of the mother cell must be conserved. For simplicity the division is assumed to take place just before the first observation of the two daughter cells at time $(i+1)\Delta t$. Cells are ordered such that cell j gives rise to cells with indices $2j$ and $2j+1$, with latent states ${}^{2j}\mathbf{X}_{t_{i+1}}^{(k)}$ and ${}^{2j+1}\mathbf{X}_{t_{i+1}}^{(k)}$, for the k^{th} particle. The conservation relationship between mother and daughter cells is enforced by requiring that ${}^{2j}\mathbf{X}_{t_{i+1}}^{(k)} + {}^{2j+1}\mathbf{X}_{t_{i+1}}^{(k)} = {}^{j}\mathbf{X}_{t_{i+1}}^{(k)}$. However, some species may be presumed to be identical between mother and daughter cells, e.g., DNA in active or inactive conformation.

After each forward simulation or division step the likelihood $P({}^{j}\mathbf{Y}_{t_{i+1}} \,|\, {}^{j}\mathbf{X}_{t_{i+1}}^{(k)}) = g({}^{j}\mathbf{Y}_{t_{i+1}}; {}^{j}\mathbf{X}_{t_{i+1}}^{(k)}, \boldsymbol{\eta})$ of each latent history is computed according to the observation function g, and used to reweight the particles. The observations of each cell are presumed independent conditional on the latent state, thus the likelihood of the complete set of observed cells is the product of the likelihoods of each cell, and the total weight of particle k is given by the product of the weights of each observed cell at that time point.

Assuming conditionally independent gamma priors $\pi(\theta_p) = \Gamma(\theta_p; \alpha_p, \beta_p)$ for each parameter θ_p, the posterior probability of the model parameters conditional on the sampled latent histories until the current time point is shifted similarly to in (Equation 9), where α_p increases by the summed number of reaction firings and β_p by the summed integrals of the (rescaled) propensity functions, over *all* newly simulated trajectories on the interval $[i\Delta t, (i+1)\Delta t]$. We define the set A_i to be the set of indices of all cells observed at any point on the interval $[0, i\Delta t]$:

$$\mathrm{A}_i = \left\{ j \,|\, {}^{j}t_1 \le i\Delta t \right\} \tag{10}$$

Let ${}^{j}r_p(i\Delta t)$ be the number of firings of reaction p in cell j at all times $t \le i\Delta t$ and ${}^{j}G_p(i\Delta t) = (1/\theta_p)\int_{{}^{j}t_1}^{\min({}^{j}t_{N_j}, i\Delta t)} a_p({}^{j}\mathbf{X}_s)ds$ be the integral of the rescaled propensity function of reaction p for cell j until time $i\Delta t$, for a particular realization of the stochastic process for cell j. With these definitions, the posterior joint density of model parameters $\boldsymbol{\theta}$ is given by:

$$P(\boldsymbol{\theta} \,|\, \mathrm{B}_i) = \prod_{p=1}^{d} \Gamma\left(\theta_p; \alpha_p + \sum_{a \in \mathrm{A}_i} {}^{a}r_p(i\Delta t), \beta_p + \sum_{a \in \mathrm{A}_i} {}^{a}G_p(i\Delta t)\right) \tag{11}$$

where the set $\mathrm{B}_i = \{{}^{a}\mathbf{X}_{[{}^{a}t_1, \min({}^{a}t_{N_a}, i\Delta t)]}\}_{a \in \mathrm{A}_i}$ gives the set of realizations of the stochastic process for all cells observed at or before time $i\Delta t$. Equation 11 provides the means to generate samples $\boldsymbol{\theta}^{(k)}$ from the probability density of model parameters conditional on a particular sampled complete genealogy B_i. The parameter samples $\boldsymbol{\theta}^{(k)}$ for each particle k are obtained by substituting the sampled trajectories for that particle into all expressions, i.e., ${}^{a}\mathbf{X}$ becomes ${}^{a}\mathbf{X}^{(k)}$, ${}^{a}r_p$ becomes ${}^{a}r_p^{(k)}$, and ${}^{a}G_p$ becomes ${}^{aa}G_p^{(k)}$. Since only the summary statistics are necessary to compute the posterior of the parameters, the full trajectories do not need to be saved, leading to a significant reduction in storage requirements. Finally, after iterating through all time points, the particles are resampled according to their weights yielding a set of K latent trajectories (if stored) and parameter sets. Thus the tree-based inference algorithm extends the single-cell-based inference algorithm (Algorithm 1) by establishing continuity between mother and daughter cells and initializing new latent trajectories for daughter cells according to the division process.

Algorithm 2: STILT algorithm

```
Data: A set of observations $\mathcal{Y} = \left\{\left({}^j\mathbf{Y}_0,\ldots,{}^j\mathbf{Y}_{N_j}\right)\right\}_{j=1}^{N_c}$ of $N_c$ cells observed at timepoints
      $\mathcal{T} = \left\{{}^1\mathbf{T},\ldots,{}^{N_c}\mathbf{T}\right\} = \left\{\left({}^jt_1,\ldots,{}^jt_{N_j}\right)\right\}_{j=1}^{N_c}$; parameter prior $\pi : \mathbb{R}^d \to \mathbb{R}_{0,+}$; observation
      function $g(\mathbf{Y};\mathbf{X},\boldsymbol{\eta}) = P(\mathbf{Y}|\mathbf{X})$; number of particles $K$; measurement time interval $\Delta t$
Result: A set of particles $\left\{\left(\left\{{}^1\mathbf{X}^{(k)},\ldots,{}^{N_c}\mathbf{X}^{(k)}\right\},\boldsymbol{\theta}^{(k)}\right)\right\}_{k=1}^{K}$ sampled from the posterior density
        $P(\left\{{}^1\mathbf{X},\ldots,{}^N\mathbf{X}\right\},\boldsymbol{\theta}|\mathcal{Y},\mathcal{T})$
initialization;
for $k$=1...$K$ do
    Sample parameter values from the prior: $\boldsymbol{\theta}^{(k)} \sim \pi(\boldsymbol{\theta})$;
    Sample initial state conditional on first observed data point: ${}^1\mathbf{X}_{t_0}^{(k)} \sim \pi(\cdot|{}^1\mathbf{Y}_{t_0})$;
    Initialize particle weight to $w_0^{(k)} := 1/K$
Generate a set of particle indices $\epsilon^{(k)} \in \{1,\ldots,K\}, k = 1,\ldots,K$ such that
    $P(\epsilon^{(k)} = a) = w_0^{(a)} / \sum_{\ell=1}^{K} w_0^{(\ell)}$;
compute maximum of all timepoints: $t_{\max} = \max(\mathcal{T})$ ;
loop over all observed timepoints;
for $i = 0 : \lceil t_{\max}/\Delta t \rceil$ do
    determine cells alive at this timepoint;
    $\boldsymbol{\sigma} = \left\{j | i\Delta t \in {}^j\mathbf{T}\right\}$ ;
    loop over particles ;
    for $k$ = 1...$K$ do
        loop over cells at current timepoint ;
        for $j \in \boldsymbol{\sigma}$ do
            Get index of current timepoint for cell $j$;
            $\ell = \text{find}({}^jt_\ell = i\Delta t)$ ;
            Compute the partial weight of particle $k$ for the $j^{th}$ cell:
                ${}^jw_i^{(k)} = P({}^j\mathbf{Y}_{{}^jt_\ell}|{}^j\mathbf{X}_{t_i}^{(k)}) = g({}^j\mathbf{Y}_{{}^jt_\ell};{}^j\mathbf{X}_{t_i}^{(k)},\boldsymbol{\eta})$;
            Generate a sample trajectory ${}^j\mathbf{X}_{[i\Delta t,(i+1)\Delta t]}^{(k)} \sim P(\cdot|{}^j\mathbf{X}_{t_i}^{(\epsilon^{(k)})},\boldsymbol{\theta}^{(\epsilon^{(k)})})$ ;
            if $(i+1)\Delta t \notin {}^j\mathbf{T}$ then
                Initialize daughter cells;
                $\left({}^{2j}\mathbf{X}_{t_{i+1}}^{(k)},{}^{2j+1}\mathbf{X}_{t_{i+1}}^{(k)}\right) \sim P(\cdot,\cdot|{}^j\mathbf{X}_{t_{i+1}}^{(k)})$ ;
            else
                Concatenate to previously sampled trajectory:
                ${}^j\mathbf{X}_{[{}^jt_1,(i+1)\Delta t]}^{(k)} := [{}^j\mathbf{X}_{[{}^jt_1,i\Delta t]}^{(\epsilon^{(k)})},{}^j\mathbf{X}_{[i\Delta t,(i+1)\Delta t]}^{(k)}]$;
        Compute the total weights for particle $k$: $w_i^{(k)} = \prod_{j\in\boldsymbol{\sigma}} {}^jw_i^{(k)}$ ;
    Generate a set of particle indices $\epsilon^{(k)} \in \{1,\ldots,K\}, k = 1,\ldots,K$ such that
        $P(\epsilon^{(k)} = a) = w_{i+1}^{(a)} / \sum_{\ell=1}^{K} w_{i+1}^{(\ell)}$;
    Sample new parameters $\boldsymbol{\theta}^{(k)}$ from the conditional density:
    $\boldsymbol{\theta}^{(k)} \sim P\left(\boldsymbol{\theta}|\left\{{}^j\mathbf{X}_{[{}^jt_1,i\Delta t]}^{(\epsilon^{(k)})}\right\}_{j=1}^{N_c}\right)$ ;
Construct a sample of $K$ particles from the posterior: $\left\{\left(\left\{{}^j\mathbf{X}_{[{}^jt_1,{}^jt_{N_j}]}^{(\epsilon^{(k)})}\right\}_{j=1}^{N_c},\boldsymbol{\theta}^{(\epsilon^{(k)})}|\mathcal{Y},\mathcal{T}\right)\right\}_{k=1}^{K}$
```

Single-Cell versus Tree-Based Inference

In the previous sections we presented two algorithms for inferring model parameters using the bootstrap particle filter. Algorithm 1 treats each cell individually while Algorithm 2 explicitly incorporates the known cellular lineage tree. As an alternative to Algorithm 2, it is also possible to fit all cells by simply discarding lineage knowledge, artificially synchronizing all cells to start at the same time point, and fitting cells in parallel. However, testing revealed that this approach quickly converges to local optima due to the inability to fit all cells simultaneously without information about their initial conditions. In contrast, Algorithm 2 benefits from exploiting the initial iterations of the algorithm with fewer cells in order to pre-converge the parameter distributions, and provides good estimates for the initial conditions of daughter cells upon division of the mother cell under the assumed division process.

The single-cell based inference performed consistently worse than the tree-based inference using STILT on synthetic data, in terms of model identification (Table S4) and parameter estimates (Figure S7). This is likely because the single-cell-based inference does not exploit the lineage structure to improve the estimation of the initial conditions (i.e., by enforcing conservation of inherited cellular material between mother and daughter cells), and because it is not obvious how to combine the inference results of individual cells in order to provide a better estimate of the overall population parameters.

Model Assumptions

Cell Division

We assume that at the time of division, each mother cell allocates its contents (mRNA and protein) randomly with equal probability to each daughter cell. Thus the number of mRNA molecules are distributed as:

$$P(M_1 = m_1, M_2 = m_2 \mid M_0 = m_0) = \frac{0.5^{m_0} m_0!}{m_1! m_2!} \delta_{m_0, m_1 + m_2} \tag{12}$$

where M_1, M_2 are the number of mRNA molecules of the two daughter cells upon division, and M_0 that of the mother cell; $\delta_{x,y}$ is the Kronecker delta. The protein contents are allocated analogously, although for numerical efficiency the binomial distribution is approximated by a normal distribution with equivalent mean and variance.

The conformation of the DNA (i.e., active or inactive) for each gene is assumed to persist from mother cell to daughter cell at division. This assumption is motivated by the observation that progeny of a cell typically resemble the ancestor cell in terms of gene activity. However, due to the stochastic nature of the model, some simulated trajectories may still switch activation state shortly after division, effectively permitting cells to also switch activation state upon division if this trajectory exhibits high likelihood. We note however that this is not an essential assumption of the inference procedure and can be easily changed for alternative scenarios; the behavior of species is easily configured in STILT.

Feedback Models

In the feedback models, the DNA activation and inactivation rates are modified by the protein abundance. For the Negative Feedback model the amount of protein modulates the rate of DNA inactivation and for the Positive Feedback model protein modulates the rate of DNA activation. We consider the case of switch-like activation/inactivation of the DNA with increasing concentration of protein. We achieve this by assuming that the propensity of DNA activation in the Positive Feedback model is given by $a_{\text{on}} = k_{\text{on}} P^2$ and of DNA inactivation by $a_{\text{off}} = k_{\text{off}} P^2$ in the Negative Feedback model.

We assume fast binding and dissociation of protein to the DNA relative to protein production and degradation, such that the protein abundance can be treated as approximately constant on the timescale of binding dynamics. With this assumption, the probability of DNA being in either the active (DNA^*) or inactive state (DNA) evolves for the Negative Feedback model according to the chemical master equation:

$$\dot{P}(\text{DNA}^*, t) = P(\text{DNA}, t) k_{\text{on}} - P(\text{DNA}^* dt, t) k_{\text{off}} P^2 \tag{13}$$

$$\dot{P}(\text{DNA}, t) = -P(\text{DNA}, t) k_{\text{on}} + P(\text{DNA}^*, t) k_{\text{off}} P^2. \tag{14}$$

Requiring that $P_{\text{SS}}(\text{DNA}^*) + P_{\text{SS}}(\text{DNA}) = 1$, the steady state solution gives:

$$P_{\text{SS}}(\text{DNA}) = \frac{P^2}{\frac{k_{\text{on}}}{k_{\text{off}}} + P^2} \tag{15}$$

$$P_{\text{SS}}(\text{DNA}^*) = \frac{\frac{k_{\text{on}}}{k_{\text{off}}}}{\frac{k_{\text{on}}}{k_{\text{off}}} + P^2}. \tag{16}$$

Thus the probability of DNA being inactive is a sigmoidally increasing function of the number of proteins. This activation function is a Hill function with coefficient 2, corresponding to cooperative binding of two Nanog molecules at the promoter/enhancer. The quantity $(k_{\text{off}}/k_{\text{on}})^{1/2}$ determines the protein abundance for which the DNA has 50% probability of being active.

Similarly, for the Positive Feedback model, the probability of the DNA states evolves as:

$$\dot{P}(\mathrm{DNA}^*,t) = P(\mathrm{DNA},t)k_{\mathrm{on}}P^2 - P(\mathrm{DNA}^*,t)k_{\mathrm{off}} \tag{17}$$

$$\dot{P}(\mathrm{DNA},t) = -P(\mathrm{DNA},t)k_{\mathrm{on}}P^2 + P(\mathrm{DNA}^*,t)k_{\mathrm{off}} \tag{18}$$

for which the steady state solution gives:

$$P_{\mathrm{SS}}(\mathrm{DNA}) = \frac{\frac{k_{\mathrm{off}}}{k_{\mathrm{on}}}}{\frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} + P^2} \tag{19}$$

$$P_{\mathrm{SS}}(\mathrm{DNA}^*) = \frac{P^2}{\frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} + P^2}, \tag{20}$$

which is a Hill function with coefficient 2 for the probability of DNA activation.

Biallelic Expression

In the synthetic datasets we consider expression dynamics of a single allele only. Thus, there are precisely four species: DNA in active and inactive conformations, mRNA and protein. In the NanogVENUS subtree modeling, the fluorescent fusion protein NanogVENUS is also expressed only in a single allele \cite{Filipzyck:2015eka}. Hence we apply the same models as for the synthetic data. However, there is also the possibility of expression in the other, unlabeled Nanog allele. Since we cannot quantify this allele, we assume that its expression is highly correlated to the observed allele, which has previously been reported for the same system (Filipczyk et al., 2013). Assuming equal proportions of observed and unobserved Nanog protein, the total amount is roughly double, which translates to a four-fold rescaling of the estimated rate constants k_{on} and k_{off} for the Positive and Negative Feedback models, respectively. The inference procedure is otherwise not affected.

Cellular Compartments

For simplification, we do not explicitly model cellular compartments such as cytoplasm or nucleus. Thus nuclear translocation effects are implicitly captured by the estimated rate constants. The NanogVENUS experiments analyzed \cite{Filipzyck:2015eka} quantify only nuclear protein. Thus the model pertains only to the expression dynamics of nuclear Nanog. The estimated protein degradation rate also captures both degradation and implicitly the dilution to the cytoplasmic compartment.

Model Specifications

Measurement Function

STILT requires the specification of an observation function that yields the likelihood of a particular observation $\mathbf{Y}_{t_k}$ given the state of the latent history $\mathbf{X}_{t_k}$ at some time t_k, this function is referred to as g above and corresponds to the measurement process. In the case of time-lapse fluorescence microscopy one typically assumes that the fluorescence intensity is proportional to the abundance of fluorophores. Assuming that the measurement process induces some small error ε, this gives the simple linear relation

$$\mathbf{Y}_{t_k} = \lambda \mathbf{X}_{t_k} + \varepsilon. \tag{21}$$

where λ is the mean fluorescence intensity per molecule. In the analysis of the synthetic lineage trees (Figure 2), no conversion between proteins and fluorescence intensity was necessary, i.e., $\lambda = 1$. We let $\varepsilon \sim N(0,\sigma^2)$ with standard deviation $\sigma = 200$ proteins. This is the same as was used for generating the noisy observations of the synthetic data.

In the analysis of NanogVENUS fluorescence lineage trees, we estimate $\lambda \approx 10^5$ proteins per unit fluorescence intensity based on comparison of mean concentration estimated from western blot and mean fluorescence intensity of unbiased NanogVENUS lineage trees (see Estimation of fluorescence intensity conversion factor). We likewise assume Gaussian measurement noise, and use the NanogVENUS lineage trees to estimate the standard deviation to be approximately $\sigma = 2 \times 10^4$ based on the small signal fluctuations. We note that the exact value of the measurement error should not bias the inference results, but rather, too small a value will lead to non-robust estimation of parameters as there is a higher risk of too many particles being discarded due to low likelihood, and too high a value leads to a poorer ability to infer model parameters as too few particles are discarded. However, the robust estimation of model parameters and apparent divergence from the prior (see Figure S3) seems to indicate an adequate choice for the magnitude of the measurement error σ.

Prior Distributions

STILT is a Bayesian inference technique and thus requires specification of prior distributions for model parameters. We utilize gamma priors distributions for each parameter, which greatly simplifies the sampling procedure (see Gamma priors). For the in silico experiments, the true model parameter were known. In this case, the priors distributions were chosen so that they i) contain the true model parameters, and ii) allow for easy visual assessment of convergence to the true model parameters.

For the investigation of the Nanog subtrees, the prior distributions were obtained as follows.

mRNA Degradation

The half-life of Nanog mRNA in mouse ES cells cultured in serum/LIF has been estimated as 6.8 hr (Ochiai et al., 2014), and 3.9-6.4 hr (Sharova et al., 2009). We thus chose $\alpha = 8$ and $\beta = 40$ for which the 95% confidence interval of the half-life is (1.9 hr, 8.0 h).

Transcription

The Nanog mRNA transcription rate was recently estimated as approximately 125 molecules/h in serum/LIF (Ochiai et al., 2014). Moreover, the number of Nanog mRNA molecules in mouse ES cells under serum/LIF conditions is approximately 300 or fewer, rarely exceeding 400 (Nair et al., 2015; Singer et al., 2014). Using the mean estimated degradation rate of 0.2 h^{-1}, and assuming DNA remains active, the expected number of mRNAs (given by k_m/g_m) would thus be approximately 625 molecules which is more than typically expected. We therefore set the prior distribution constants to be $\alpha = 10, \beta = 0.1$ for which the 95% confidence interval of the transcription rate becomes $[47.95, 170.85]$ h^{-1}, and the expected number of transcripts is approximately 500. This number is somewhat reduced by the fact that the DNA is typically not persistently active, and by the cell division process which reduces the mRNA count by a factor of approximately 2.

Protein Degradation

The half-life of Nanog has been reported as approximately 2.1 ± 0.8 h (Sokolik et al., 2015). Nanog's half-life was also measured in the analyzed dataset and found to be closer to 5 hr \cite{Filipzyck:2015eka}. We therefore conservatively set the prior distribution constants to be $\alpha = 2, \beta = 4$ such that the 95% confidence interval of the half-life is approximately $0.5 - 11.5$ h.

Translation

The translation rate of Nanog is not well characterized. However, the estimated mean number of Nanog molecules per cell is approximately 350,000, the mean degradation rate approximately 0.2 h^{-1}, and mean number of mRNA molecules roughly 200. In the deterministic limit, the expected number of proteins is given by $\langle p \rangle = (k_m/g_m)\langle m \rangle$, where $\langle p \rangle$ and $\langle m \rangle$ denote the mean protein and mRNA counts, respectively. Substituting our estimates, we obtain approximately 350 $mRNA^{-1}$ h^{-1}. We thus chose $\alpha = 3, \beta = 0.01$ such that the mean (std.) translation rate is 300 ± 173 $mRNA^{-1}$ h^{-1}. We note that this is in rough agreement with the mean estimated translation rate of 478 $mRNA^{-1}$ h^{-1} over all analyzed genes in mouse fibroblasts (maximum estimated rate of 1000 $mRNA^{-1}$ h^{-1}) (Schwanhäusser et al., 2011).

DNA Activation and Inactivation

The rate of DNA activation and inactivation for Nanog is not well studied. By inspecting the analyzed NanogVENUS subtrees, we surmise that periods of rapid fluorescence intensity likely correspond to periods of DNA activity, and periods of decline to DNA inactivity. Thus activation and inactivation presumably proceeds with expected waiting time on the order of hours and not days or longer. Consistent with this, Sokolik et al. estimated active/inactive switching times to be approximately 3.8 ± 1.2 h (Sokolik et al., 2015). Since the autoregulation models studied differ in the form of their activation/inactivation rates, it was necessary to choose priors for each separately. For each of the three models, the prior distributions for k_{on} and k_{off} were chosen such that the waiting times for activation/inactivation were on the order hours, and such that trees simulated with these parameters produced reasonable dynamics, i.e., observed DNA state switching, and approximately correct order of magnitude for number of proteins.

Initial Conditions

The initial state of the founder cell of the cellular lineage trees is unknown. In the considered models DNA and mRNA are entirely latent, while protein is observed with noise. For both the in silico experiments and the NanogVENUS subtrees, we initialized each cell's DNA state to be active or inactive with 50% probability, and to have an initial mRNA count uniformly sampled from $[0, 50]$ molecules. The protein copy number was sampled from a Gaussian distribution centered on the first observation, with standard deviation specified by the measurement function (see Measurement function).

Number of Particles

All implementations of the bootstrap particle filter require specification of the number of particles to use for approximating the latent history and posterior parameter distributions. The accuracy of the approximation improves in the Monte Carlo sense as the number of particles is increased. However, the incurred computational overhead increases proportionally. We used 7×10^5 particles for inference of the synthetic lineage trees, 10^5 for each synthetic cell for the single-cell-based algorithm, and 10^6 particles for the NanogVENUS subtrees. We determined the number of particles to use based on robustness of convergence of the parameter posteriors, and run-time. Inference typically completed on a multicore machine in approximately 10h for 10^6 particles for a single subtree.

Implementation

STILT (see Algorithm 2) was implemented using MATLAB 2015. It includes code for importing SBML models and fast, parallel stochastic forward simulations for the system state using MATLAB's parallel computing toolbox.

Model Definition via SBML

Our implementation supports the import of biochemical network models from SBML using libSBML 5.12.0 (Bornstein et al., 2008), but can also be specified directly in MATLAB. Species, reactions and their parameters are translated into a stoichiometric matrix and vectorized MATLAB functions for computing reaction propensities (see Chemical reaction networks).

Simulation

The stochastic simulation code was implemented using explicit, adaptive τ-leaping (Cao et al., 2006, 2007), which generates approximate samples from the exact stochastic process. In general, τ-leaping approximates the Markov jump process by a Poisson process with the same expected number of reactions firing for time intervals where the reaction propensities remain relatively constant.

Such an approximation is generally necessary when the system becomes stiff, i.e., when there exist reactions with widely varying timescales such as is the case for the protein production and degradation compared to DNA activation and inactivation. As for (Cao et al., 2007) our implementation distinguishes between critical and non-critical reactions (based on current educt availability, the parameter N_c is set to 10) and performs explicit τ -leaping (with parameter $\varepsilon = 0.03$) for non-critical reactions with error bounding implemented for first, second and third order reactions. The simulation code was implemented completely vectorized and provides the approximated integrated reaction propensities (for mass action propensities only) and the number of reactions firing, which are required for inference (see Inference of latent history and model parameters). The forward simulation code can be further accelerated by converting the entire function or only the Poisson random number generation to C code, which lead to significant speedup for the studied systems.

Data Structure

Measurement data are specified using a generic MATLAB structure containing measurement times, cell number and an indicator for censoring (e.g., for inaccurate or missing data) as well as measurements and their respective measurement errors. Field names of measurements are automatically matched to SBML species. Parameter priors, model specifications (e.g., behavior on cell division), compilation behavior and other user configurations are provided via an options structure.

Model Evaluation

Marginal Likelihoods and Bayes Factors

The particle filtering approach presented above can be used for performing model comparison via Bayes factors, i.e., by computing $P(M_1 \mid \mathbf{Y}_{0:N}) / P(M_2 \mid \mathbf{Y}_{0:N})$, the ratio of the posterior probabilities of Model 1 (M_1) to Model 2 (M_2) for any two models. As before, we denote the series of observations at times $t_0, \ldots, t_N$ by $\mathbf{Y}_{0:N} = (\mathbf{Y}_{t_0}, \ldots, \mathbf{Y}_{t_N})$. Using Bayes' law, one can reformulate the marginal posterior probability of a model M as:

$$P(M \mid \mathbf{Y}_{0:N}) = \frac{P(\mathbf{Y}_{0:N} \mid M) P(M)}{P(\mathbf{Y}_{0:N})} \tag{22}$$

Following Wilkinson et al. (Wilkinson, 2011), p. 294, we can approximate the marginal likelihood of the model $P(\mathbf{Y}_{0:N} \mid M)$ using the sampled particles at each iteration i. First, the distribution of the observed data at time t_{i+1} depends only observations up to t_i: $P(\mathbf{Y}_{t_{i+1}} \mid \mathbf{Y}_{0:N}, M) = P(\mathbf{Y}_{t_{i+1}} \mid \mathbf{Y}_{0:i}, M)$. Moreover, this probability is approximated by the expectation of the likelihood, or weights $w^{(k)}$, of the particles:

$$\begin{aligned} P(\mathbf{Y}_{t_{i+1}} \mid \mathbf{Y}_{0:i}, M) &= \int P(\mathbf{Y}_{t_{i+1}} \mid \mathbf{X}_{t_{i+1}}) P(\mathbf{X}_{t_{i+1}} \mid \mathbf{Y}_{0:i}, M) d\mathbf{X}_{t_{i+1}} \\ &\approx \frac{1}{K} \sum_{k=1}^{K} \underbrace{P\left(\mathbf{Y}_{t_{i+1}} \mid \mathbf{X}_{t_{i+1}}^{(k)}\right)}_{w_{i+1}^{(k)}} \end{aligned} \tag{23}$$

where the $\mathbf{X}_{t_{i+1}}^{(k)}$ are sampled (via the particle filter) from the marginal posterior up to time t_{i+1} given by $P(\mathbf{X}_{t_{i+1}} \mid \mathbf{Y}_{0:i}, M)$. This is nothing more than a Monte Carlo approximation of the integral, which provides an unbiased approximation of $P(\mathbf{Y}_{t_{i+1}} \mid \mathbf{Y}_{0:i}, M)$ with variance decreasing as K^{-1} (Doucet and Johansen, 2009).

Next, since the distribution of each observation depends only on previous observations, the marginal probability of the entire set of observations $P(\mathbf{Y}_{0:N} \mid M)$ is given by the product:

$$P(\mathbf{Y}_{0:N} \mid M) = P(\mathbf{Y}_{t_0}) \prod_{i=0}^{N} P\left(\mathbf{Y}_{t_{i+1}} \mid \mathbf{Y}_{0:i}, M\right). \tag{24}$$

Assuming a priori equally likely models, the factor of $P(M)$ in Equation 22 cancels between the two models and the Bayes factor reduces to the ratio of marginal likelihoods. In the analysis presented in this work we primarily utilize log Bayes factors and marginal log likelihoods due to their superior numerical performance.

Goodness-of-Fit Test

To assess the extent to which a particular model agrees with an observed dataset, we developed a simple goodness-of-fit (GOF) test. The GOF test utilizes an estimate for model parameters obtained from the particle filter to generate many synthetic datasets, which are then compared against the measured data. Specifically, we use the assumed model to generate many synthetic lineage trees of the same number of generations as the observed data using the median posterior parameter estimate of each parameter. For each newly simulated dataset, we approximate its log likelihood conditioned on the parameter set that was used to generate that data. The conditional log likelihood (CLL) is approximated again via a particle filter, where the parameters are fixed. This conditional particle filter only samples from the latent history of all state variables while keeping the parameters fixed, and is essentially the same as in Algorithm 2 omitting the parameter resampling step.

To compensate for the fact that the simulated datasets and the measured dataset do not necessarily contain the same number of transitions, we normalize the estimated CLL of each simulation by the number of simulated transitions (i.e., between measurement time points). We likewise normalize the CLL of the actual data by the number of transitions (subtracting censored observations), to obtain the average CLL per transition. Without this compensation the CLL is always decreasing with the number of transitions since the log likelihood is never greater than zero, which could potentially bias the CLL depending on the random lifetime of each simulated cell.

In all applications of the GOF we used 300 samples to approximate the null distribution of the CLL, and 500 particles per sample to approximate the CLL. The GOF test approximates the null distribution of the CLL, i.e., the distribution of CLL values yielded by the particle filter when the parameters and model utilized are known to be true. We compute the CLL of the actual dataset using the same parameter values and compare it with the null distribution of the CLL. If it lies within this distribution, then with high probability the dataset could have been produced by the chosen mechanistic model and parameter set, and the model cannot be rejected. Conversely, if the CLL of the observed dataset lies outside the null distribution, the model and parameters are unlikely to have produced this dataset. Thus, we define three categories of model agreement with the null distribution: reject ($p < 0.02$), marginal ($p < 0.05$), and accept, otherwise. Empirical p values are estimated using the empirical cumulative distribution function of the estimated CLL.

Time-Lapse Fluorescence Microscopy Data

Pre-processing

We obtained quantified time-lapse fluorescence microscopy movies of NanogVENUS in mouse embryonic stem cells from the dataset of Filipczyk et al. \cite{Filipzyck:2015eka} and converted fluorescence intensities to protein numbers (see Measurement function).

Since the time-lapse fluorescence microscopy quantification introduces error due to variability in the cellular (nuclear) segmentations, background correction, etc., we performed a data cleaning step prior to analysis. We censored measurements for all automatically segmented cells that could not be manually verified, e.g., if the cells were too densely packed or overlapping to be reliably quantified. We further censored very large jumps (the top 5% of absolute change in intensity) in the quantified intensity of individual cells, which result from either contamination due to microscopic debris, errors in cell segmentation, and in some cases jumps in the intensity at the last time point before cell division which presumably arise due to a sudden change in the cellular morphology preceding division that leads to a large overlap of cytoplasmic and nuclear volumes (see Figure S8 for examples). Censored measurements affect the inference by rendering all simulated particles equally likely at that iteration; the algorithm otherwise proceeds as normal.

After cleaning and selecting datasets, we obtained a total of 7 quantified cellular genealogies from 3 different experiments. To improve computational efficiency, and for comparison with the synthetic data, we subdivided these large trees into smaller subtrees each containing 7 cells with no overlap between subtrees (see Figure 3A), thus obtaining 4 subtrees from the first experiment, 8 subtrees from 3 different parent trees of the second experiment, and 1 subtree from each of 3 parent trees of the third experiment; in total 15 subtrees were used for further analysis (see Figure S2).

Estimation of Fluorescence Intensity Conversion Factor

To estimate the absolute number of NanogVENUS molecules per cell, we performed western blots experiments on 10% polyacrylamide gels. We compared western Blots with a known quantity of NanogGFP single knockin fusion ESCs and with different quantities of recombinant GFP (Catalogue number: 632373, Clonetech, CA, USA). Both NanogGFP and GFP proteins were detected using an anti-GFP primary antibody consisting of two monoclonal clones (Catalogue number: 1181460001, Roche, Mannheim, Germany). Western blot band intensities were quantified by using the Gel Analyzer tool in FIJI to gate on protein lanes and quantify band intensities over background. We found that the relationship between the GFP quantities x and the corresponding intensity y is best described by a sigmoid function:

$$y(x) = \left(\frac{\lambda x^n}{K^n + x^n}\right)\varepsilon \tag{25}$$

The model parameters λ_j, K_j, n_j were obtained by local optimization using multiple restarts initialized according to Latin-hypercube sampling. The exponent n determines the shape of the sigmoid function, K sets the inflection point, λ is the maximum of the curve and ε is a log-normally distributed error term with expectation 1 and standard deviation σ, as is suggested for western blot data (Kreutz et al., 2007). We compared this model against linear models both with and without intercept and found it to be superior according to both the Bayesian Information Criterion and coefficient of variation between replicates. We solve Equation 25 for x to obtain

$$x = \frac{K_j}{\left(\frac{\lambda_j}{y_j} - 1\right)^{\frac{1}{n_j}}}. \tag{26}$$

The total quantity of protein x is related to the cellular average P_j as

$$x = P_j c w, \tag{27}$$

where c is the number of loaded cells and w is the molecular weight for the protein of interest. Thus we determined the number of proteins P_j per cell from the sample intensity y_j of each western blot replicate j by first computing x from the observed intensity y according to Equation 25, and then substituting into Equation 27.

As P_j is a combination of uncertain variables, we obtained error bars for each P_j individually by applying standard error propagation to account for uncertainties in the number of cells c (we assume a standard deviation of 10%) and uncertainties in the model (estimated via the standard deviation σ of our noise model). However, we find that the uncertainties for each individual replicate P_j are always smaller than the inter-replicate standard deviations by a factor of 0.3 or smaller. Therefore, we only consider the standard deviation across replicates, as this is the dominant source of uncertainty in our procedure. Finally, we determined the fold-change between NanogVENUS and NanogGFP from three western blots. Uncertainty of protein abundance over replicates was estimated by simple error propagation. All above experiments and analysis were performed in triplicate.

The resulting estimate is of approximately $350,000 \pm 72,000$ molecules of NanogVENUS expressed in each cell. Using the distribution of NanogVENUS fluorescence intensities over unsorted mESCs, we obtain a median intensity of approximately 3.5, from which we determine the calibration factor of NanogVENUS fluorescence intensity to NanogVENUS molecules count to be approximately 100,000.

Experimental Validation of Negative Feedback Model

Exogenous NanogVENUS Expression Compartments

Using the wild-type cell line which expresses no NanogVENUS, we obtained the NanogVENUS intensity distribution due only to autofluorescence (Figure S4B), from which we deduce the 0.95 quantile of NanogVENUS autofluorescence, denoted I_{neg}^{NV}. We then used the expression distribution of a cell line which expresses NanogVENUS at one endogenous allele (NV cell line, see Exogenous Nanog construct) to derive the 0.95 quantile of unperturbed NanogVENUS expression, denoted I_0^{NV}.

We define the NanogVENUS "No Exogenous" compartment as NanogVENUS fluorescence intensities that are below I_{neg}^{NV}. Since NanogVENUS is only expressed on one allele of the NV cell line, the total quantity of Nanog in the cell is approximately twice this amount. Based on this we define the 1x overexpression (OE) compartment to be intensities that are above the No Exogenous compartment but below 200% of the normal level I_0^{NV}. The 2x OE compartment has NanogVENUS intensities I^{NV} between 2-4 times normal, and the 3x OE compartment between 4-6 times normal. Cells with higher NanogVENUS intensities fall into the "very high" intensity compartment see table below).

k	NanogVENUS compartment	Intensity range (I_k^{NV})	NanogVENUS median expression (I_{med}^{NV}, % of I_0^{NV})
1	No Exogenous	$(-\infty, I_{\text{neg}}^{NV})$	9.55%
2	1x OE	$[I_{\text{neg}}^{NV}, 2\times I_0^{NV})$	56.56%
3	2x OE	$[2,4)\times I_0^{NV}$	263.87%
4	3x OE	$[4,6)\times I_0^{NV}$	482.38%
5	very high	$[6\times I_0^{NV}, \infty)$	1254.41%

Downregulation of Endogenous Nanog Levels

To investigate negative feedback, we utilize the NK cell line (see Exogenous Nanog construct), and compute the expression levels of endogenous NanogKATUSHKA for different levels of exogenous transgenic NanogVENUS expression. We first obtain the autofluorescence intensity distribution on the NanogKATUSHKA channel using NV cells which express no KATUSHKA, from which we estimate the 0.95 quantile of intensity of the KATUSHKA negative compartment, denoted I_{neg}^{NK} (Figure S4D). We first normalize NanogKATUSHKA fluorescence relative to background by subtracting I_{neg}^{NK} from the measured intensities. We then compute the median fold-change, for each overexpression compartment k, of normalized expression of NanogKATUSHKA relative to that of the No Exogenous compartment (see Figure 4E; see table above). Denoting the NanogKATUSHKA expression of a cell with index j by ${}^jI^{NK}$, and its NanogVENUS expression by ${}^jI^{NV}$, the median fold-change for compartment k is given by:

$$FC_k^{NK} = \frac{median_{\{j:{}^jI^{NV}\in I_k^{NV}\}}\left({}^jI^{NK} - I_{neg}^{NK}\right)}{median_{\{j:{}^jI^{NV}\in I_1^{NV}\}}\left({}^jI^{NK} - I_{neg}^{NK}\right)}. \tag{28}$$

Comparison of Experimental Replicates

The NanogVENUS overexpression experiment described above was performed three times. To compensate for batch effects, the distributions in each experiment were normalized relative to the first experiment. Specifically, for each replicate, the fluorescence intensity of endogenous NanogKATUSHKA was scaled linearly so that the median intensity of cells matches to the median intensity of the first experiment. The same NanogKATUSHKA background level I_{neg}^{NK} was used for each of the three replicates (see Downregulation of endogenous Nanog levels).

In Silico Perturbation Experiment

We replicate the experimental setup by extending the Negative Feedback model to include exogenous Nanog (P_{ex}), such that the propensity of DNA inactivation becomes

$$a_{\text{off}} = k_{\text{off}}(P_{\text{en}} + P_{\text{ex}})^2. \tag{29}$$

Parameter	Value	Units
k_{on}	0.6854	h^{-1}
k_{off}	5.690×10^{-11}	h^{-1} protein^{-2}
k_m	110.3	h^{-1}
g_m	0.226	h^{-1} mRNA^{-1}
k_p	619.5	h^{-1} mRNA^{-1}
g_p	0.220	h^{-1} protein^{-1}

From the time-lapse fluorescence microscopy movies of NanogVENUS subtrees we obtain estimates of posterior distributions of parameters for the Negative Feedback model (Table S6). We compute the median of the posterior for each parameter and subtree, and then the mean of the medians for each parameter over the subtrees (see table below). We then perform in silico perturbation experiments using these mean parameter values and various levels of exogenous Nanog.

To mimic the experimental setup, we sampled intensity values I_{ex}^{NV} directly from the measured distributions of exogenous NanogVENUS for each overexpression compartment separately. We convert the sampled intensities into a specific number of molecules by computing the overexpression relative to wild-type NV cells. Since the fluorescent reporter is expressed only on one allele, a 100% increase of NanogVENUS corresponds to an approximately 50% increase in total Nanog levels. We assume approximately 2×10^5 NanogVENUS molecules in a cell with no exogenous perturbation (see Nanog Bandplots in Mendeley Data archive). We thus compute the sampled amount of exogenous NanogVENUS molecules as $P_{ex} = (1/2)(I_{ex}^{NV}/I_0^{NV})(2\times10^5)$.

For each sampled quantity of exogenous NanogVENUS, we generated 50 synthetic lineage trees of 5 generations each. The founder cell of each lineage tree had DNA initially active, between 0 and 150 mRNA molecules (uniformly sampled), and between 10^5 and 2×10^5 Nanog molecules (uniformly sampled). The exogenous Nanog levels were held fixed at the sampled value for the duration of the simulation.

Comparison to Simulations

The distribution of endogenous Nanog following perturbation was computed for each overexpression compartment after 46h of simulated time. The fold-change of endogenous Nanog expression relative to expression with no perturbation was computed analogously to Equation 28:

$$^{j}FC_k^{sim} = \frac{^{j}N_k^{sim}}{median\left(^{j}N_1^{sim}\right)} \tag{30}$$

where $^{j}N_k^{sim}$ denotes the number of endogenous Nanog molecules of cell j, 46 hr after the perturbation corresponding to compartment k.

In Figure 4E, we plot the distribution of the fold-change of simulated cells with respect to the No Exogenous compartment as a box-and-whiskers plot (median shown as red line). We compare this against the median (mean, s.e.m. from 3 experimental replicates) fold-change computed for the experimental data, see Equation 28. The comparisons against each experimental replicate individually are shown in Figure S4E.

DATA AND SOFTWARE AVAILABILITY

Software

STILT is available for download at http://www.imsb.ethz.ch/research/claassen/Software/stilt—stochastic-inference-on-lineage-trees.html.

Data Resources

The NanogVENUS fluorescence lineage tree quantifications analyzed are available upon request.

The sampled trajectories (Bandplots) for Nanog subtrees (protein and mRNA), and the distributions produced by the goodness-of-fit test for the Nanog subtrees are available for download from Mendeley Data at https://data.mendeley.com/datasets/wx6s4mj7s8/2.