DISS. ETH NO. 22356

INTERROGATING THE SINGLE CELL: COMPUTATIONAL AND EXPERIMENTAL METHODS FOR OPTIMAL LIVE CELL EXPERIMENTS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES OF ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

MICHAEL PETER UNGER

Dipl.-Ing., Graz University of Technology, Austria
born on 12.01.1985
citizen of Austria

accepted on the recommendation of

Prof. Dr. Heinz Koeppl (examiner)
Prof. Dr. Matthias Peter (co-examiner)
Prof. Dr. Serge Pelet (Assistant Professor SNF) (co-examiner)

2014

The only way to have a friend is to be one.

— Ralph Waldo Emerson

Dedicated to the loving memory of Stefan Nikolaus Graf.

1928 – 2013

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to Heinz and Matthias. Heinz, I would like to thank you for providing me with this interdisciplinary position, and for having the confidence to give me the invaluable freedom to explore, learn about and combine multiple topics of research. I am truly grateful for your constant support and supervision and for giving me numerous opportunities to travel and interact with other researchers. Matthias, I would like to thank you for welcoming and integrating me to your lab and for allowing me to pursue my project. Thank you for your continued trust, support and valuable advice.

I would like to thank Serge and Sung Sik. You both have been amazing teachers and and I always enjoyed the time working with you.

Further, I would like to thank all past and present members of the Koeppl and Peter Groups, especially Christoph, Preetam, and Björn for our inspiring collaborations. Both places have been tremendous and I always enjoyed my time there. I also thank the staff of both the Automatic Control Lab, especially Martine, Alain and Markus, and the Institute of Biochemistry, especially Barbara, Nico and Toni.

I wish to express my appreciation to Sebastian Maerkl, for being in my thesis committee, and for letting me visit his lab at EPFL Lausanne, Switzerland, to learn about microfluidics, to Dan Larson, for hosting me in his lab at NIH, Bethesda, MD, USA, and to Gustavo Stolovitzky and everyone involved in organizing the HPN-DREAM Breast Cancer Network Inference Challenge for allowing us to be part of it.

Thank you Anne, Rado and Björn, for your help and friendship in and outside the lab.

I would like to express my heart-felt gratitude to my family, my parents Hildegard and Gerhard, my brother Markus and his wonderful family, and my grandparents for all their love and support.

Finally, I would like to thank Maren for being with me through all the little highs and lows, and reminding me about important things in life.

CONTENTS

ACRONYMS

AC Adenocarcinoma

AUPR Area Under Precision-Recall curve

BCM Belief Confusion Metric

BLUF Blue-Utilizing Flavin

CCEM Correct Class Enrichment Metric

CFP Cyan Fluorescent Protein

ChIP Chromatin Immunoprecipitation

CLE Chemical Langevin Equation

CME Chemical Master Equation

COPD Chronic Obstructive Pulmonary Disease

CTMC Continuous Time Markov Chain

CHX Cyclohexamide

DIC Differential Interference Contrast

DPP Dynamic Prior Propagation

DREAM Dialogue for Reverse Engineering Assessments and Methods

ELISA Enzyme Linked Immunosorbent Assay

ER Estrogen Receptor

FCS Fluorescence Correlation Spectroscopy

FOV Field Of View

FIM Fisher Information Matrix

FISH Fluorescence *In Situ* Hybridization

FLIP Fluorescence Loss In Photobleaching

FRAP Fluorescence Recovery After Photobleaching

FRET Fluorescence Resonance Energy Transfer

FP Fluorescent Protein

FSP Finite State Projection

GAL4DBD GAL4 DNA-Binding Domain

GEF GTPase Exchange Factor

GEV GAL4DBD.ER.VP16

GFP Green Fluorescent Protein

HPN Heritage Provider Network

HOG High Osmolarity Glycerol

HU Hydroxyurea

IFFL Incoherent Feed Forward Loop

KL Kullback-Leibler

LASSO Least Absolute Shrinkage and Selection Operator

LatA Latrunculin A

LNA Linear Noise Approximation

LoC Lab-on-a-Chip

LOV Light-, Oxygen- or Voltage-sensing

LC Lung Cancer

MAP Maximum A Posteriori

MAPK Mitogen-Activated Protein Kinase

MCMC Markov Chain Monte Carlo

MSE Mean Squared Error

MEMS Microelectromechanical Systems

μTAS Miniaturized Total Analysis Systems

MGF Moment Generating Function

MS Multiple Sclerosis

MSD MS Diagnostic

MCS Multiple Cloning Site

NaCl Natriumchlorid

NES Nuclear Export Signal

NGS Next Generation Sequencing

NLS Nuclear Localization Signal

ODE Ordinary Differential Equation

OED Optimal Experimental Design

OID Optimal Input Design

PCR Polymerase Chain Reaction

PDMS Polydimethylsiloxane

PFS Perfect Focus System

POC Point-Of-Care

pSTL1 STL1 Promoter

PWM Pulse Width Modulation

qV quadruple Venus

RFP Red Fluorescent Protein

RMA Robust Multi-array Average

sbv IMPROVER Systems Biology Verification combined with Industrial Methodology for Process Verification

SCC Squamous Cell Carcinoma

SCV Squared Coefficient of Variation

SDE Stochastic Differential Equation

SMC Sequential Monte Carlo

SSA Stochastic Simulation Algorithm

TNF Tumor Necrosis Factor

TRAIL TNF-Related Apoptosis-Induced Ligand

TCA Trichlor Acid

TF Transcription Factor

UTR Untranslated Region

VP Virus Protein

WT Wild Type

YFP Yellow Fluorescent Protein

SUMMARY

Advances in quantitative single cell experimental techniques and mathematical modeling have enabled insights to cellular processes that have not been possible before. Experimental protocols, however, often remain intended for intuitive interpretation, and resulting data may not be optimally informative for the reverse engineering of kinetic parameters, molecular states or network topologies. The central goal of this thesis was to develop computational and experimental methods for the optimized investigation of single cell dynamics.

We introduce model-based Optimal Experimental Design (OED) methods to automatically determine temporal concentration profiles to perturb the extracellular environment and sufficiently excite the biological system under study. Based on information-theoretic arguments, perturbation sequences are designed to maximize the expected information gain between consecutive experiments. A Bayesian modeling framework allows to incorporate prior knowledge from preceding experiments to iteratively refine the model of interest. We introduce microfluidic approaches for synthesizing such concentration sequences, using the concept of Pulse Width Modulation (PWM) of liquid flows, during live cell fluorescence microscopy experiments. A valve-off-chip approach allows for an easy integration into existing experimental settings, while a dedicated valve-on-chip PDMS platform combines the PWM approach with single cell traps to capture and track individual live cells. In two case studies, we demonstrate the combination of dynamic single cell fluorescence microscopy recordings and a Bayesian inference scheme. In the first we applied an artificial gene expression system induced by the hormone β-estradiol to express a short-lived fluorescent protein. A dedicated calibration curve allows us to map fluorescent intensities from microscopy recordings to absolute copy numbers of fluorescent proteins within single cells. In a second case study, we investigate the effect of different temporal concentration profiles on the information gain between consecutive inference runs for a model system of Hog1 induced gene expression. We demonstrate that complex sequences lead to a significantly larger information gain than input perturbations of lower complexity.

In the next part of the thesis we use a stochastic model to formalize and develop new hypothesis about the sensing and decoding of chemical gradients, and test these experimentally. Using a combination of quantitative microscopy, microfluidic tools, stochastic modeling and genetic engineering, we could identify a molecular mechanism of gradient sensing in yeast, and derive general design principles of cell-cell communication systems.

Finally, we introduce our contribution to a scientific challenge for the identification of diagnostic signatures for multiple diseases.

ZUSAMMENFASSUNG

Fortschritte in quantitativen experimentellen Methoden mit Auflösungen von einzelnen Zellen, in Kombination mit mathematischen Modellen, haben Einsichten in zelluläre Prozesse ermöglicht, die davor nicht denkbar waren. Experimentelle Protokolle sind jedoch oft für direkte Interpretation der Daten ausgelegt und eignen sich nur bedingt für die Rekonstruktion von kinetischen Parametern, molekularen Konfigurationen oder Modell-Topologien. Das Ziel dieser Arbeit war es mathematische und experimentelle Methoden zu entwickeln, um die Dynamik einzelner Zellen zu studieren.

Modell-basierte Methoden zur Optimierung von Experimenten erlauben es, Zeitprofile von chemischen Konzentrationen zu bestimmen, um molekulare Systeme optimal anzuregen. Diese Profile werden entworfen, um den Informationsgewinn, beschrieben durch informationstheoretische Maße, zwischen zwei aufeinanderfolgenden Experimenten zu optimieren. Basierend auf Bayesscher Statistik können wir Vorkenntnisse aus vorangegangenen Experimenten einbinden und iterativ verbessern. Wir verwenden Pulsweitenmodulation (PWM), um Profile von Konzentrationen für Fluoreszenz-Mikroskopie Experimente zu synthetisieren. Eine Implementierung verwendet dazu externe Ventile und kann gut in bestehende Vorrichtungen integriert werden. Eine PDMS *microfluidic* Plattform mit integrierten Ventilen kombiniert das PWM Konzept mit Vorrichtungen, um individuelle Zellen zu fixieren. In zwei Beispielen zeigen wir, wie Mikroskopiedaten von angeregten dynamischen Systemen für Bayessche Inferenz verwendet werden können. Im ersten verwenden wir ein künstliches Genexpressions System, basierend auf dem Hormon β-Estradiol, um fluoreszierende Proteine mit verkürzter Halbwertszeit zu exprimieren. Eine spezifische Kalibrierungskurve erlaubt es, von Fluoreszenzsignalen auf die Anzahl von fluoreszierenden Molekülen pro Zelle zu schließen. In dem zweiten Beispiel untersuchen wir den Effekt von verschiedenen Konzentrationsprofilen auf den Informationsgewinn zwischen aufeinanderfolgenden Inferenz Iterationen an einem Modell von Hog1 induzierter Genexpression. Wir zeigen, wie komplexere Profile im Vergleich zu einfacheren Profilen einen deutlich höheren Informationsgewinn erzeugen.

Im nächsten Teil dieser Arbeit entwickeln wir ein stochastisches Modell, welches uns bei der Formulierung von Hypothesen über das Entschlüsseln von chemischen Gradienten in zellulären Systemen hilft. In einer Kombination aus quantitativer Mikroskopie, *microfluidics*, stochastischen Modellen und genetischen Modifikationen konnten wir einen molekularen Mechanismus in Hefe entschlüsseln und generelle Prinzipien von Zell-Zell Kommunikationssystemen ableiten.

Abschliessend stellen wir unseren Beitrag zu einem wissenschaftlichen Wettbewerb vor, bei dem Diagnostische Signaturen etabliert werden mussten.

Part I

PRELIMINARIES

1 INTRODUCTION

Quantitative biology is a data-driven science. It aims to understand complex molecular mechanisms that govern dynamic cellular behavior by combining quantitative experimental techniques, mathematical modeling, statistical analysis and computational tools.

Over the last decades, the development of effective experimental methods reached a precision where ever more measurements become available at a single cell level. Unlike more traditional approaches that worked on the average of a whole population of cells, the advent of high-throughput single cell techniques allows to monitor the individual behavior of cells, and how their states or responses to stimuli might differ. Awareness of cell heterogeneity [7, 140] is becoming increasingly important to understand regulatory mechanisms on the cell population level or in developmental biology, in the diagnosis and treatment of diseases, such as autoimmune diseases or cancer [43, 203], or in investigating the development of drug resistance [57]. An increase in cell-to-cell variability, was also associated with aging, as a consequence of cellular degeneration and DNA damage [16].

The origin of cell heterogeneity can have various causes [181, 140, 209], but even for genetically identical cells, stochastic processes such as gene expression lead to fluctuations, or noise (e.g. in protein levels) between multiple cells. Gene expression noise in particular could be caused by transcription bursts in mRNA production, the ratio of protein lifetimes to the intervals between protein production bursts, or the propagation of fluctuations due to the stochastic expression of transcription factors and other upstream components themselves [59]. On a cell population level stochastic heterogeneity can increase robustness [174] and act beneficial for the population survival in response to stress (e.g upon osmotic stress [176]), the commitment to irreversible differentiation processes (e.g. in TRAIL induced apoptosis [211]), or by phenotypic switching in unpredictable fluctuating environments (e.g. like *Escherichia coli* upon antibiotic treatment [17], or as investigated using a *S. cerevisiae* model system [4]), to name but a few.

Various experimental techniques are available that can capture data on a single cell level. Population snapshot methods typically allow for a high throughput, but can not capture temporal dynamics within single cells, as any correlation between individual cells is necessarily lost. These methods include flow- [258] or mass-cytometry [19, 21] – in contrast to fluorescent probes used in flow-cytometry, mass-cytometry uses isotope antibody labels to overcome limitations due to fluorescence spectra overlaps – but also techniques such as Fluorescence *In Situ* Hybridization (FISH) [165] that allow to detect and localize specific copies of DNA or RNA within fixed cells. Recently, a whole new set of technologies, such as single cell

RNA-seq or ChIP-seq, emerged, which are built around Next Generation Sequencing (NGS) tools [202]. They allow to sequence DNA-based or DNA-convertible readouts and enable unprecedented analysis of genetic, epigenetic, transcriptional or proteomic targets and their specific interactions.

Variants of time-lapse light microscopy, like epifluorescence microscopy, confocal microscopy, or Fluorescence Correlation Spectroscopy (FCS) on the other hand, are especially well suited to follow the temporal and spatial dynamics of molecular processes within individual live cells over time. Plenty of different imaging and labeling strategies are available, and allow to tailor a solution to the respective needs. Currently available methods and their applications are discussed in more detail in Section 2.2.

Microscopy is particularly suited to study temporal dynamics as it allows for an easy combination of data acquisition and alteration of cellular states using light [133, 252, 20, 227, 220, 127, 157], or through manipulating extracellular environments using microfluidic techniques [96, 224, 232, 231, 91, 90, 259].

Photochemical approaches became widely available with the advent of synthetic biology and offer unprecedented spatial and temporal resolution of perturbing cellular processes in live cells. Tools for photo-controlling processes such as protein function, transcription, degradation, or translocation, enzymatic activity or chromatin modification can be adapted and used in a variety of systems. A conformational change of the photoactuator-fused protein can then be induced by shining light of a particular wavelength to directly alter protein activity, to control protein-protein interactions, or to release an agonist or antagonist of a particular protein. Natural light-sensitive proteins, in particular flavoproteins like the Light-, Oxygen- or Voltage-sensing (LOV) proteins, Blue-Utilizing Flavin (BLUF), and the plant light-sensitive cryptochrome (CRY2), or other photoreceptors such as plant phytochromes (PHYs) have been re-engineered to be used as photoactuators. Other approaches use a hybrid between genetic modifications of proteins and exogenous photoactive synthetic molecules. [73]

The development of microfluidics, a technology of manipulating fluids on a micro-liter scale, had a significant impact on many fields related to diagnostics or biomedical research. Microfluidic devices, often referred to as Miniaturized Total Analysis Systems (μTAS) or Lab-on-a-Chip (LoC) technologies are often built using lithography technologies that were first developed by the semiconductor industry, and later sucessfully applied to Microelectromechanical Systems (MEMS). Especially the use of Polydimethylsiloxane (PDMS) in combination with *soft lithography*, which allowed a relatively easy, flexible and cheap production of microfluidic chips, led to the success and wide acclaim of microfluidics. Due to its optical properties, its gas permeability, and elasticity – which enables the possibility of manufacturing valves on a chip – PDMS microfluidics caused a downright revolution of possible chip designs and assays. [191] A more detailed description of microfluidics, current applications and downsides are discussed in Section 2.1.

As ever more experimental techniques with unprecedented accuracy and throughput become available, scientists can start addressing increasingly complex questions, and aim to understand complex biological networks involving many components on a quantitative level. Therefore, processing and interpretation of generated data with the aim of understanding complex dynamic behavior needs to be aided by mathematical modeling and statistical analysis. Combining experimental techniques and computational methods is fundamental to answer questions at the core of systems biology, as it aims to understand [116]:

- The structure of intracellular (e.g. signaling networks, metabolic networks, gene expression mechanisms, and their interactions) and multicellular systems (e.g. in developmental biology or the reaction of cell populations).
- The systems dynamics, and how cells behave over time in various conditions.
- The control and modulation of systems to influence a cellular state and potentially prevent or reverse diseases.
- The design and modification of systems for specific tasks, guided by design principles and simulations.

Building predictive models is essential to all of the above listed points. Many models of intracellular processes neglect the spatial location of molecules in a cellular compartment, and assume a small well-stirred reservoir to describe the dynamical process of interest using a set of biochemical reactions. Modeling approaches then differ fundamentally in what chemical kinetics formalism is used to describe the reaction networks.

More traditional approaches have been centered around *classical chemical kinetics* and use a deterministic approach, described by a set of Ordinary Differential Equations (ODEs). These approaches can be good approximations in cases where reactants are available in high abundance. Reactants are typically measured in concentrations, and on a continuous scale. Given a fully parameterized model and the starting conditions, the model behavior is completely predetermined. Efficient computational tools are available to numerically integrate even large reaction networks, and to fit models to experimental data using various optimization techniques, including local methods, such as least-squares fitting and gradient-based methods, or global methods, such as simulated annealing or evolutionary algorithms [158, 13]. However, deterministic approaches fail to capture the inherent heterogeneity and intrinsic stochasticity that many biological systems exhibit [246].

Analysis of data generated with modern single cell experimental techniques thus requires a comprehensive modeling framework based on probability theory, to fully cover the dynamics and variability of the underlying biological processes. Particularly processes where reactants of low

copy numbers are involved (e.g. gene expression mechanisms; see Section 2.3), require the application of *stochastic chemical kinetics*, where reactions happen at random when a specific combination of discrete reactants interacts with each other [194, 246]. Modeling approaches using stochastic processes, such as Continuous Time Markov Chains (CTMCs) [78, 10] provide a suitable solution. The state of a CTMC reflects the numbers of molecules of individual reaction-components available, and changes as a reaction happens, with the probability of a reaction to happen depending only on the current state. The probabilities that the system is in any possible state as a function of time are described by the Chemical Master Equation (CME), but analysis through the CME turns out to be infeasible for any but the simplest models. Another way of analyzing a CTMC model is to investigate simulated trajectories, much like they would be the result of an experiment. Exact realizations of trajectories, or sample paths over a specified time frame, starting from a given initial state, can be drawn using the Stochastic Simulation Algorithm (SSA) [77]. Probabilistic properties of the underlying process can then be computed using Monte Carlo methods, with the accuracy increased as the number of sample paths used becomes larger. In practice however, computing a reasonable number of sample paths can be computationally infeasible, and approximations or alternative approaches become essential.

Approximate methods of simulating sample paths include τ-leaping methods [81, 82, 183, 225, 36], which use a simulation time step that is long enough to cover multiple reactions, and assume that changes within a single time step are negligible. Other approaches, such as the Linear Noise Approximation (LNA) [60, 95] or the Chemical Langevin Equation (CLE) [79, 80] aim at approximating the Markov process by a diffusion process, described by a set of Stochastic Differential Equations (SDEs). The Finite State Projection (FSP) offers a method to directly solve or approximate the CME by truncating the possible state space [159].

To extract mathematical models and their properties from experimental data, inverse problems – the reverse-engineering, or inference of network topologies, model parameters or molecular states of an underlying system – have to be solved.

The inference of mechanistic stochastic model topologies [254, 170, 137] is still largely limited by the scope and precision of quantitative measurement techniques. The estimation of kinetic model parameters, and the reconstruction of molecular states, however, has already been an active topic of research [84, 94, 258, 259, 137]. Contrary to deterministic models, no direct distance function is available, and the numerical analysis of the likelihood function is again computationally demanding [246]. Various inference approaches are based on Monte Carlo methods [259], or approximations of the exact stochastic model [190].

Even though unprecedented experimental and computational methods are available, individual experiments are still often designed for direct interpretation, or limited by constraints on machine or personnel time,

financial resources, or ethics, with the resulting data serving only as a compromise for learning computational models. Optimal design principles [39] can be employed to chose the most informative of possible experimental parameters, such as measurement time points or inducer concentrations, to address specific problems such as reducing estimation uncertainties or non-identifiabilities [136, 137]. Experimental design strategies are often centered around information criteria, such as the Fisher information, and have been applied to problems such as the inference of model parameters [18] or network topologies [215, 12, 33, 208]. Increasingly, Optimal Experimental Design (OED) strategies are developed for stochastic chemical kinetics [120, 256, 189]. Combination of novel experimental techniques with tools of perturbing cellular states, such as microfluidics, are well suited for the design of optimal temporal perturbation sequences, or Optimal Input Design (OID), to maximally excite a molecular pathway under study [151, 162, 256].

The development of novel experimental techniques, as well as the increased availability of computational resources triggered the rise of systems biology. Biology turned more into a quantitative science, and interactions with other disciplines, such as mathematics, statistics and physics were strengthened. Significant advances in the reconstruction of cellular interaction networks, or in learning predictive models were made. Fundamental questions about the validity of these approaches, and the resulting models, however, remained open. Initiatives, such as the Dialogue for Reverse Engineering Assessments and Methods (DREAM)[1] [217, 218, 178, 179, 145] have been established with the aim of a fair comparison of the strengths and weaknesses of methods, the assessment of how well respective models describe underlying biological systems, and to strengthen collaboration among scientists. Systems Biology Verification combined with Industrial Methodology for Process Verification (sbv IMPROVER)[2] [154, 155, 223], a follow up initiative, aims to apply a related approach using crowd sourcing to benchmark scientific methods in an industrial context.

1.1 AIM OF THE THESIS

This thesis aims at realizing the inherent potential of mathematical methods to gain unprecedented insights into biological processes, by developing and aligning computational and experimental methods for the optimized investigation of single cell dynamics.

This includes the development of computational methods to design perturbation sequences to generate experimental data, optimized for specific modeling tasks, such as the reverse-engineering of kinetic parameters, molecular states or network topologies, the stochastic modeling of cellular processes, the development of fluorescence microscopy assays, including

1 http://www.the-dream-project.org
2 http://www.sbvimprover.com

the optimization of fluorescent reporter constructs, and the design and fabrication of novel microfluidic devices.

1.2 SUMMARY AND CONTRIBUTIONS

This thesis is split into five parts. In the remaining Chapter 2 of the introductory Part i, I introduce mainly experimental concepts that are essential to the subsequent parts. I will give primers on various topics, list current advances and open questions, and discuss contributions to the respective fields that are not, or only partly reflected in the remaining manuscripts. These primers focus on microfluidics, and especially the subclass of PDMS devices, concepts of fluorescence microscopy and biological systems to study gene expression networks, which were used as model systems in the following parts.

Parts ii - iv are thematically cohesive blocks, and reproduce manuscripts which contain significant contributions of this thesis.

The major part of this thesis is collected in Part ii. I start with Chapter 3, where I introduce the concept of Pulse Width Modulation (PWM) of liquid flows to synthesize temporal perturbation profiles of media concentration during live cell microscopy experiments. PWM is a well established concept in electrical engineering, that allows to bring many of the advantages of digital electronics, such as its robustness to noise (e.g. pressure fluctuations), to manipulating liquids at the micro- or milliliter scale. The function of the device is demonstrated using the HOG pathway by maintaining the Mitogen-Activated Protein Kinase (MAPK) Hog1 activation in budding yeast cells by ramping of extracellular Natriumchlorid (NaCl) concentration.

Chapters 4 and 5 focus on the design of optimal input stimuli (i.e. temporal perturbation sequences) for the identification of stochastic reaction dynamics. In the Chapter 4, we propose a framework to generate constrained sequences, and compute expected information theoretic measures, such as the Kullback-Leibler (KL) divergence to assess the predicted capability of these sequences to excite a molecular pathway under study. We analytically proof the optimal input sequence for a simple Birth/Death model, given complete observations, and provide simulation studies for cases where only noisy and incomplete measurements are available. In Chapter 5, the design of optimal sequences is extended as a variational problem, which we aim to solve with a numerical approach, using a gradient based algorithm based on stochastic approximation. Simulation case studies demonstrate that the generalized posterior variance of estimated model parameters could be reduced by orders of magnitude.

Chapter 6 discusses a Bayesian inference framework, Dynamic Prior Propagation (DPP), that incorporates population heterogeneity. The framework allows the inference of kinetic model parameters, the reconstruction of inaccessible molecular states, the computation of Bayes factors for model selection tasks, and the dissection of noise into intrinsic, extrinsic

and technical contributions, using time-lapse, single cell measurements. We applied the inference scheme to an artificially controlled gene expression system in the yeast *Saccharomyces cerevisiae*. Using a combination of fluorescence time-lapse microscopy, and a microfluidic flow chamber, we recorded data of transcription factor relocation, and resulting expression reporter intensity. In order to follow the transcription of reporter mRNA closer, we increased the degradation of the reporter constructs by adding a destabilizing sequence. We developed a novel approach of mapping fluorescence intensities to total protein abundances, using a calibration curve, fit to measurements taken from reference strains.

In Chapter 7, I aim to combine the optimal design of input perturbations with the DPP inference framework to study the gene expression network, controlled by the HOG pathway in the yeast *S. cerevisiae*. I designed a novel microfluidic chip that combines the PWM approach of synthesizing temporal profiles of chemical concentrations with single cell traps. This multi-layer, valve-on-chip design significantly reduces the volume of media consumed during a specific time-frame, and in combination with the single cell traps, allows for an increased experiment duration, and enhanced imaging quality. I show how the reduction of parameter uncertainties for iterative inference cycles depends on the specific datasets used.

In Part iii, we applied methods for single cell analysis and developed a stochastic computational model to investigate cell-cell communication and understand its underlying mechanisms. Cell-cell communication, in processes such as neutrophil chemotaxis, or chemotrophic growth in neurons or yeast, requires cells to sense and decode shallow chemical gradients. We chose to investigate the pheromone response of budding yeast as a model system, as the network structure is highly conserved across single and multicellular eukaryotes. Combining fluorescence single cell microscopy, a microfluidic platform to generate gradients of α-factor, and chemical or genetic perturbations, we could observe the dynamics of how the polarity sites of yeast cells orient towards the direction of the extracellular gradient. The observations were collected in a stochastic model to identify essential components of a generalizable regulatory network that governs polarity establishment and the sensing of directional cues. We found that a spatial, double positive feedback loop between the polarity module and the gradient receptor is sufficient to reliably decode the gradient information. A fast acting feedback loop establishes cell polarity in an initially random direction, while the second, slower loop aims at orienting and stabilizing the polarity site towards the gradient source. A tight control of the feedback system, and partial activation of the polarity module through expression-mediated sequestration of the polarity activator turned out to be essential to establish a mobile polarity site. Through sequential assembly and disassembly cycles, the polarity site takes repeated decisions in which directions to move. As these decisions are more likely to happen towards the stronger gradient side, the position of the polarity site will slowly converge towards the gradient's maximum.

I discuss our contributions to scientific crowd-sourcing challenges in Part iv. These challenges are part of an ongoing effort to assess and benchmark computational methods for the reverse-engineering of biological processes in systems biology. Our approach to the sbv IMPROVER Diagnostic Signature Challenge, where we were ranked as the 3rd best performing team, out of more than 50 submissions, is described in Chapter 9. We developed an approach, based on L_1-regularized logistic regression to classify unlabeled clinical samples based on transcriptomics data, and establish diagnostic signatures in the disease areas of Psoriasis, Multiple Sclerosis (MS), Chronic Obstructive Pulmonary Disease (COPD), and Lung Cancer (LC). After the challenge was closed, we analyzed similarities and differences in top-scoring approaches, investigated *wisdom of the crowd* effects and discussed persistent difficulties. These findings can be found in a manuscript [223], not reproduced in this thesis. For the Heritage Provider Network (HPN) DREAM breast cancer inference challenge[3], we were involved in the organization, development and formulation, and led the development of the *in silico* part of the challenge. A manuscript describing the overall challenge and its outcome in great detail is currently in preparation, but is not reproduced in this thesis.

Finally, I draw conclusions in Part v. I discuss how current limitations could be addressed, and which directions future work should take. Further, I highlight possible applications of the methods developed within the frame of the thesis.

1.3 RELATED PUBLICATIONS

Related to Part II

- M. Unger, S. S. Lee, M. Peter, and H. Koeppl, *Pulse width modulation of liquid flows: towards dynamic control of cell microenvironments*, in 15th International Conference on Miniaturized Systems for Chemistry and Life Sciences, 2011, pp. 1567-9.
- P. Nandy, M. Unger, C. Zechner, and H. Koeppl, *Optimal perturbations for the identification of stochastic reaction dynamics*, in 16th IFAC Symposium on System Identification, 2012, pp. 686-91.
- C. Zechner, P. Nandy, M. Unger, and H. Koeppl, *Optimal variational perturbations for the inference of stochastic reaction dynamics*, in 51st IEEE Conference on Decision and Control, 2012, pp. 5336-41.
- C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, *Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings*, Nature Methods, vol. 11, no. 2, pp. 197-202, doi. 10.1038/nmeth.2794, Jan. 2014.

3 https://www.synapse.org/#!Synapse:syn1720047/wiki/

- M. Unger, C. Zechner, S. S. Lee, S. Pelet, M. Peter, and H. Koeppl, *Optimizing single-cell recordings for the inference of transcription dynamics*, In Preparation

Related to Part III

- B. Hegemann, M. Unger, S. S. Lee, I. Stoffel-Studer, J. van den Heuvel, S. Pelet, H. Koeppl, and M. Peter, *A cellular system for spatial signal decoding*, Submitted

Related to Part IV

- P. Nandy, M. Unger, C. Zechner, K. K. Dey, and H. Koeppl, *Learning diagnostic signatures from microarray data using L1-regularized logistic regression*, Systems Biomedicine, vol. 1, no. 4, p. e25271, 2013.
- A. L. Tarca, M. Lauria, M. Unger, E. Bilal, S. Boue, K. K. Dey, J. Hoeng, H. Koeppl, F. Martin, P. Meyer, P. Nandy, R. Norel, M. Peitsch, J. J. Rice, R. Romero, G. Stolovitzky, M. Talikka, Y. Xiang, C. Zechner, *Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge.*, Bioinformatics, vol. 29, no. 22, pp. 2892-9, Nov. 2013. *(This manuscript is not reproduced in this thesis.)*
- HPN-DREAM Consortium, *The HPN-DREAM Network Inference Challenge*, In Preparation *(This manuscript is not reproduced in this thesis.)*

Not Discussed in this Thesis

- N. Hiroi, M. Klann, K. Iba, P. de Heras Ciechomski, S. Yamashita, A. Tabira, T. Okuhara, T. Kubojima, Y. Okada, K. Oka, R. Mange, M. Unger, A. Funahashi, and H. Koeppl, *From microscopy data to in silico environments for in vivo-oriented simulations*, EURASIP Journal on Bioinformatics and Systems Biology, vol. 7, Jan. 2012. *(This manuscript is not reproduced in this thesis.)*

2 PRIMERS

2.1 MICROFLUIDICS

Microfluidics, as a broad term, describes an entire field that recently emerged and deals with the engineered handling and manipulation of liquids using devices with channels on the micrometer scale [212]. The application of microfluidic devices had a significant impact on many fields, such as biomedical research, drug development or diagnostics [244, 191].

µTAS approaches for diagnostics have been especially valuable in developing countries, where only very limited resources are available for the processing of clinical samples [251, 143]. A microfluidic Point-Of-Care (POC) approach of an Enzyme Linked Immunosorbent Assay (ELISA), for example, has been implemented and successfully applied to diagnose HIV and syphilis in patients in Rwanda [42]. The test performance was comparable to reference benchtop assays, but came at a significant lower cost, while having reduced complexity in its application and interpretation. Other approaches aim at reducing cost and complexity of diagnostic µTAS devices by making use of materials such as paper or wax [147].

A new sub-class of microfluidics, so called *organ-on-a-chip* devices, aim to replicate *in vivo* organ function on a µTAS device. Such devices could have significant impact on the pharmaceutical industry by enabling the testing of compounds or drugs before, or even instead of animal trials. Recent developments of *organ-on-a-chip* devices include lung [103], blood-vessels [228] or cancer models [261, 22].

Applications in biomedical research span a wide range of different functions, including numerous designs to study cell dynamics. Approaches are often centered around switching between different input flows [96, 91, 231], establishing gradient profiles [132], or the capturing of single cells [46].

2.1.1 *PDMS Microfluidics*

Microfluidics emerged as an indispensable part in numerous biomedical research laboratories. This was largely due to the adoption of Polydimethylsiloxane (PDMS) as a material that allowed a relatively easy and cost effective method for designing and manufacturing of custom microfluidic devices.

PDMS is a soft polymer material that combines several advantageous properties for the use in microfluidics [243, 168]. These include:

- Optical properties allow to use chips for microscopy experiments, as it is transparent over a wide spectrum from visible to near ultraviolet light.
- Low cost of essential equipment, and the easy design, fabrication and setup of experiments with custom chips.
- Low toxicity, the surface chemistry and it's gas permeability make it suitable for cell culture applications.
- Reversible and irreversible bonding of PDMS to materials such as glass or PDMS itself.
- The elasticity of PDMS allows for easy manufacturing using replica-molding, and the implementation of elements such as on-chip valves [233].

On the other hand, there are also some downsides in PDMS based microfluidics, including the release of uncrosslinked oligomers [184] or the difficulties to scale up the production of chips. One of the most critical ones for the use in biomedical research, however, remains the following:

- PDMS has been shown to absorb small hydrophobic molecules, which can significantly change the solution concentration [226, 184]. An effect on the experimental outcome can be especially critical when studying cellular dynamics. No generally applicable solution to this has been identified so far.

The fabrication of PDMS based microfluidic devices is commonly done by replica molding. The drawing of a new design is typically printed on transparent film or glass, which is then used as a mask for photolithography. The required printing resolution depends on the size and the desired precision of the smallest structures in the drawing. For each layer, photoresist is applied onto a silicon wafer, shined with UV light through the photomask to polymerize exposed regions, developed and washed. The resulting wafer with the remaining photoresist structures serves as a mold for casting PDMS. The mixed polymer PDMS is poured over the master mold, baked and peeled of. Holes for reservoirs or tubing connections are punched. To close the channel structures, the surface of the PDMS layers can be oxidized using a plasma or UV lamp and irreversibly fused to glass or another PDMS layer [168].

Detailed protocols of manufacturing the wafer moldes, as well as for the PDMS chip fabrication, for the example of the microfluidic device introduced in Chapter 6, can be found in the Appendix Chapter A.

2.2 MICROSCOPY

Variants of light microscopy have proven to be an indispensable tool for studying cell dynamics, since they allow to follow cellular processes in

individual cells over time. While cell snapshot techniques, like FISH or various immunostaining approaches, where fixed cells are imaged, can give information about the molecular variability within a cell population, time-lapse live cell microscopy variants have the additional benefit of capturing temporal correlations within single cells.

2.2.1 *Concepts of Live Cell Microscopy*

The selection of a suitable microscopy technique [139, 216] has to be carefully matched to the individual requirements of imaging a specific cellular process and cell type. One needs to consider the brightness of the signal and the required detection sensitivity, the spatial and temporal resolution and experiment duration needed to properly capture the process dynamics, the shape and viability of cells, or the number of different wavelengths that have to be sampled. Typically no individual microscopy technique will fulfill all criteria of a specific experiment the best, and compromises have to be made. In the following I will briefly discuss some of the essential microscopy concepts to study cellular dynamics in live cells.

2.2.1.1 *Fluorescence Microscopy*

The major innovation that revolutionized fluorescence microscopy [135], and with it research in cell biology in general, was the introduction of the Green Fluorescent Protein (GFP) [38, 229]. GFP is a protein isolated from the jellyfish *Aequoria victoria*, that can be integrated and expressed in genetically modifiable organisms and fused to other proteins of interest to visualize their abundance and position [238]. A wide variety of Fluorescent Proteins (FPs) [201] with different characteristics are currently available and allow the imaging of multiple targets simultaneously.

Fluorophores, molecules with fluorescent properties like FPs in general, can be excited with a specific wavelength, and emit light of a slightly shifted wavelength. The difference between these two, called the Stokes shift, allows to filter out the excitation light, such that only light emitted from the fluorophores is detected.

Fluorescence microscopy is often complimented with traditional bright-field, or Differential Interference Contrast (DIC) images to capture important information about the cell location, shape and vitality.

2.2.1.2 *Variants of Fluorescence Microscopy*

Numerous variants of microscopy that make use of specific properties of fluorescent probes emerged. Such methods are particularly useful for studying the dynamics of cellular processes, such as the movement and interaction of targets [139, 216].

Fluorescence Recovery After Photobleaching (FRAP) or Fluorescence Loss In Photobleaching (FLIP) are techniques that use the precise control over the region of illumination that confocal microscopy offers to specifically

photobleach a defined region within a cell. Both methods are used to study the mobility of FP-tagged proteins. In FRAP experiments, fluorescent molecules are irreversibly photobleached using a laser beam. The subsequent recovery of fluorescence signal through the diffusion of non-bleached FPs into the photobleached area is recorded and gives a measure of the mobile fraction of FPs and the diffusion constant. Instead of observing fluorescence recovery, FLIP monitors the loss of fluorescence in a defined area that is actively bleached, while also recording images of the whole cell. By observing the loss of fluorescence intensity in both areas, FLIP can thus give information about connected regions within a cell. [242, 139]

Two methods that allow to measure protein-protein interactions using fluorescence microscopy are Fluorescence Resonance Energy Transfer (FRET) and Fluorescence Correlation Spectroscopy (FCS).

In FRET microscopy, the proximity of two proteins, can be visualized by tagging the involved proteins with spectrally overlapping fluorophores, a FRET pair, such as a Cyan Fluorescent Protein (CFP) and a Yellow Fluorescent Protein (YFP). If a *donor* label of the first protein, e.g. CFP, is excited, the energy absorbed will be transfered to, and thus excite, an *acceptor* label, e.g. YFP, of the second protein in very close proximity. Since the efficiency of the energy transfer decays with the sixth power of the distance between the pair, FRET is an ideal tool to study protein interactions. [139, 109]

Since confocal microscopy and two-photon excitation allow to focus on very small sampling volumes, the popularity of FCS improved steadily. FCS allows to measure the fluctuations of fluorophores moving in and out of the sampling volume, and thus gives a measure of the average number of molecules within the volume, as well as the diffusion speed. By correlating multiple species, FCS allows to study protein-protein interactions. [62]

2.2.1.3 *Fluorescent Labeling Strategies*

Not only the developments in microscopy technologies, but also advances of fluorescent probes and labeling strategies [49] are essential components for extending the possibilities of imaging dynamic processes in single cells.

N or C-terminal tagging of proteins with FPs in the yeast is typically done by chromosomal integration of a Polymerase Chain Reaction (PCR) amplified cassette with overlapping homologous sequences [198, 15, 117], and toolboxes containing various vectors and optimized proteins exist [108, 131]. A wide variety of FPs with different excitation and emission spectra, size and structure are available [167], and allow to label and monitor a few proteins, in practice typically around 3-4, at the same time. To monitor transient events, like transcription events, closer, efforts have been made to reduce the half-lives of FPs, using the N-end rule protein degradation pathway [85, 236, 237, 102]. Figure 1 shows a comparison of time series recordings between a stable Venus FP, and a destabilized version.

Other approaches of imaging transcription events aim at visualizing RNA directly. Therefore, a RNA binding protein from bacteriophage, like

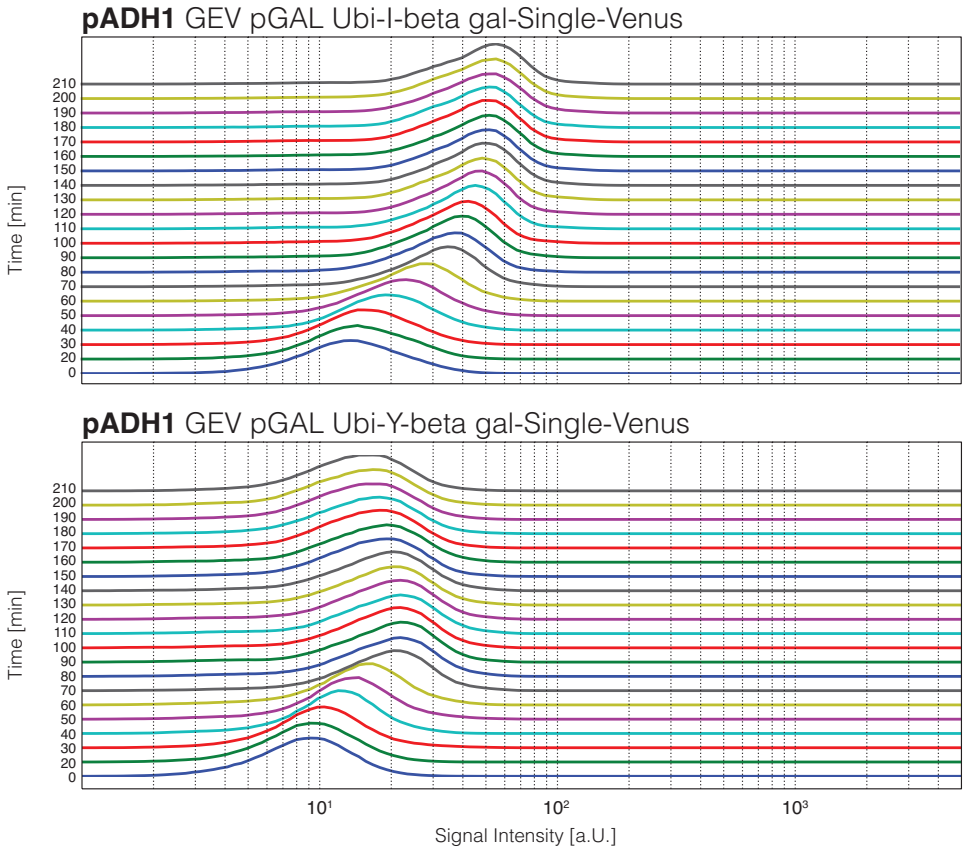

Figure 1: Time series comparison of a stable (Ubi-I-beta) and a destabilized (Ubi-Y-beta) version of a Venus FP under the control of a GAL promoter in the yeast *S. cerevisiae*. Cells were induced with 100 nM β-Estradiol for 60 min before blocking translation using Cyclohexamide and measured using flow cytometry.

MS2 or PP7, that recognizes specific RNA motifs, can be fused to a FP and the specific RNA motif can be genetically incorporated to the 3′ or 5′ UTR of the RNA. The RNA stem-loop motifs can be integrated repeatedly to achieve single-transcript resolution by trapping multiple FP-fused proteins. [23, 128, 129] Such systems have been applied to study the transcription dynamics in yeast [129] and mammalian systems [127].

Recently, novel reporters have been developed that can translate kinase activity into nuclear/cytoplasmic relocation events, using a phosphoregulated Nuclear Localization Signal (NLS) and Nuclear Export Signal (NES) [185].

Instead of fluorescent proteins, synthetic fluorophores, such as fluorescent dyes, can be used to label target proteins. Such labeling approaches often rely on tag-mediated systems, like the SNAP [113], CLIP [74], and HELO tags [141]. [100, 49] Efforts have been made to overcome the limited accessibility of exogenously supplied dyes to yeast cells [150, 213].

2.3 INDUCIBLE GENE EXPRESSION SYSTEMS

Gene expression mechanisms lend themselves well to study the variability of molecular mechanisms, due to the low abundance of many components involved, particularly the single copies of specific genes. Such systems are particularly useful model systems for the reverse-engineering of network structures or kinetic parameters from input-output data. Therefore, inducible gene expression systems are of specific interest, as they allow, for instance using microfluidic devices, to control the induction time and dose. Using fluorescent reporters, resulting mRNA or proteins can be imaged and quantified. To study transcription initiation processes, a FP can be put under control of a promoter that is driven by the Transcription Factor (TF), activated by the respective inducer.

2.3.1 *The GEV System*

The GAL4DBD.ER.VP16 (GEV) [142, 149] system is built around the chimeric TF GEV, a fusion of the GAL4 DNA-Binding Domain (GAL4DBD), with the hormone-binding domain of the human Estrogen Receptor (ER) [142], and the transcription-activating domain of the herpes simplex Virus Protein (VP) VP16 [192]. GEV can be expressed in the yeast *S. cerevisiae*, where its inactive, cytoplasmic form associates with the Hsp90 chaperone complex. By adding the hormone β-estradiol, Hsp90 disassociates from the complex, active GEV translocates to the nucleus, and acts as a TF for genes under a GAL1 promoter. The range of β-estradiol concentration for a gradual increase of gene expression can be shifted by changing the amount of GEV present in a cell. Figure 2 shows example dose response plots, acquired by flow cytometry, for two endogenous promoters, expressing GEV in different strengths. An application of the GEV system can be found in Chapter 6.

2.3.2 *The HOG Pathway*

The High Osmolarity Glycerol (HOG) pathway [193] in the yeast *S. cerevisiae* triggers essential cellular response and adaptation mechanisms upon high external osmolarity. The core of the pathway is a MAPK cascade [40] with two different branches that activate either Ssk2/Ssk2 (these proteins are homologous and functionally redundant) or Ste11 and converge on the MAPK kinase Pbs2, which is the specific activator that double phosphorylates Hog1. Upon activation, the MAPK Hog1 translocates to the nucleus, where it drives transcriptional responses. These responses include the expression of proteins under control of promoters such as STL1 (activated by TFs Hot1 and Sko1), CTT1 or ALD3 (TFs Msn2 and Msn4) or HSP12 (TFs Msn2, Msn4, Hot1) [37]. Hog1 induced gene expression has been shown to exhibit bimodal characteristics [176] (see also Figure 3) and is thus ide-

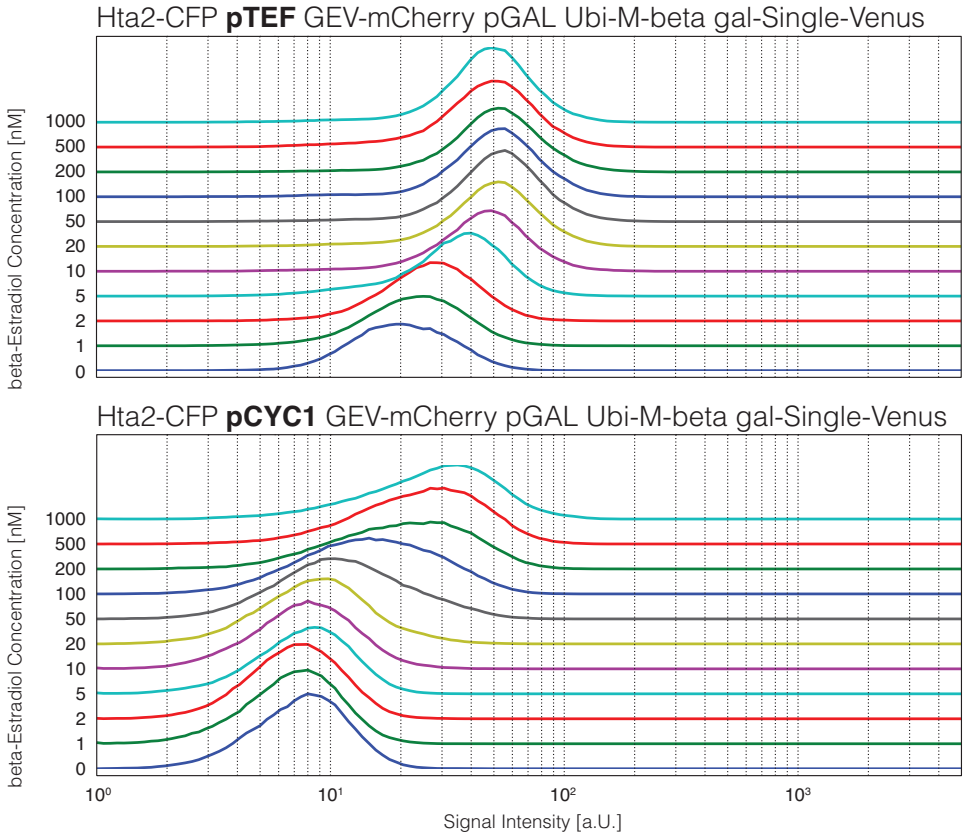

Figure 2: Dose responses of β-estradiol induced expression of a Venus yellow fluorescent protein under a GAL promoter in the yeast *S. cerevisiae*, using the GEV system. Changing the amount of the TF GEV present by placing it under different endogenous promoters (e.g. pTEF or pCYC1), the induction range of the system can be shifted. Cells were induced with the respective concentration of β-estradiol for 60 min before blocking translation using Cyclohexamide, incubated for 60 min, and measured using flow cytometry.

ally suited for analysis using stochastic models [258]. The HOG pathway is used as a model system in Chapter 7.

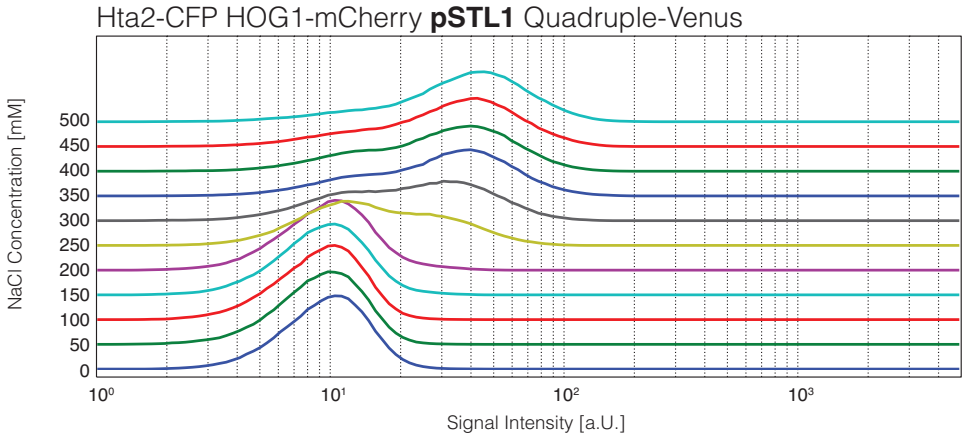

Figure 3: Dose response of Hog1 mediated gene expression of a quadruple Venus fluorescent reporter, driven by a STL1 promoter in the yeast *S. cerevisiae*. Cells were induced with the respective concentration of NaCl for 45 min before blocking translation using Cyclohexamide, incubated for 90 min, and measured using flow cytometry.

Part II

INTERROGATING THE SINGLE CELL: COMPUTATIONAL AND EXPERIMENTAL METHODS FOR OPTIMAL LIVE CELL EXPERIMENTS

3

MANUSCRIPT 1

REFERENCE

M. Unger, S. S. Lee, M. Peter, and H. Koeppl, *Pulse Width Modulation of Liquid Flows: Towards Dynamic Control of Cell Microenvironments*, in 15th International Conference on Miniaturized Systems for Chemistry and Life Sciences, 2011, pp. 1567-9.

AUTHOR CONTRIBUTIONS

MU, SSL, MP and HK participated in designing the project. MU and SSL implemented the method. MU performed measurements. MU, SSL and HK wrote the manuscript.

PULSE WIDTH MODULATION OF LIQUID FLOWS: TOWARDS DYNAMIC CONTROL OF CELL MICROENVIRONMENTS

M. Unger[1,2], S.S. Lee[2,3], M. Peter[2,3], and H. Koeppl[2,3]

ABSTRACT

Advanced methods for live cell analysis help to understand fundamental processes within a cell. We propose a method to generate specific input stimuli that are essentially needed to get insights to dynamic cellular behavior. By Pulse Width Modulation of liquid flows, we offer a fast and reliable way to generate temporal profiles of media concentration that can be combined with other microfluidic devices. Its functionality is demonstrated in experiments of salt concentration ramping, which is of high interest and biological relevance for maintaining activation of the MAPK Hog1 in the HOG pathway.

KEYWORDS

Pulse Width Modulation, Concentration Ramping, Signaling Pathway

INTRODUCTION

Pulse Width Modulation (PWM) is a well-established concept in electrical engineering [101]. We make use of this concept and transform it to a method for generating temporal profiles of media concentration within cell microenvironments. This shows to be of vital importance, as recent works [161, 176, 262] have reported the need of advanced input stimuli for the understanding of intracellular dynamics of the HOG pathway in the yeast, *Saccharomyces Cerevisiae*.

THEORY

A desired output value of electric voltage or current can be realized by averaging over a time series of fast switches between the ON and OFF states of a power supply. The switching ratio between these two states, and thus the width of the individual pulses, defines the realized output value. We now take the concept of PWM from electrical engineering and apply it to the modulation of liquid flows on a microfluidic scale. This allows a precise dilution of a desired medium concentration [6]. We extend the

1 BISON Group, Automatic Control Laboratory, D-ITET, ETH Zurich, Switzerland
2 Institute of Biochemistry, D-BIOL, ETH Zurich, Switzerland
3 Competence Center for Systems Physiology and Metabolic Diseases, Switzerland

application of PWM to generate dynamic profiles of media concentration over time. Examples of concentration ramping will be given at a later stage.

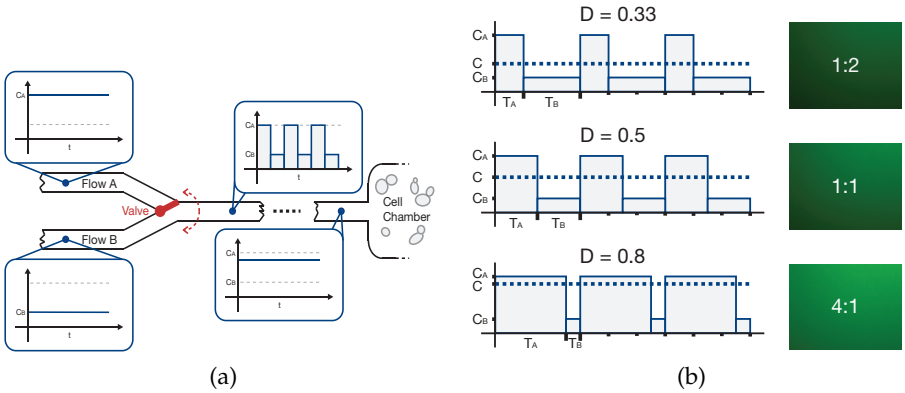

Figure 4: (a) Schematic illustration of the Pulse Width Modulation (PWM) setup, as used for perturbing cell microenvironments. The switching between two different input flows is done by a computer controlled solenoid valve. This is connected to the cell chamber by PTFE tubing, serving as a diffusion channel to filter out individual PWM packages and create the desired media concentration.
(b) *(left)* Specific concentrations C of media are generated in the cell chamber by adapting the PWM duty cycle D (Eq. (1)), determining the On/Off switching ratio between the two input concentrations C_A and C_B. *(right)* Experimental validation of static dilution by fluorescence microscopy. The intensity of fluorescent FITC dye represents the specific concentration of the medium in the cell chamber. The numbers within the image give the ratio of T_A/T_B.

Consider an output stream, made up of switching between two input streams of different concentrations C_A and C_B of a fluid with equal and constant flow rates. As the input streams are selected mutually exclusive, the output flow rate remains constant and equal to the input flows. Thus we can define liquid packages of equal volume transported sequentially in time T. The average amount of concentration C after diffusion (low-pass filtering) within each depends on the fraction of the total package volume consisting of media with concentration C_A and C_B. The ratio of the time the input stream with concentration C_A is active to the total package time $T = T_A + T_B$,

$$D = \frac{T_A}{T} \tag{1}$$

is referred to as duty cycle D. By determining D, and thus defining the ratio of C_A and C_B within a package, a desired output concentration C can be diluted:

$$C = \frac{D \cdot T \cdot C_A + (1 - D) \cdot T \cdot C_B}{T} \tag{2}$$

We can now extend this concept of media dilution further to dynamic dilution. By computing a new value for the duty cycle D for each liquid

package of time T, we can control the output concentration profile over time:

$$C[t] = D[t] \cdot C_A + (1 - D[t]) \cdot C_B \tag{3}$$

If, for example, the duty cycle is steadily increased up to $D = 1$, the concentration ramps to the maximum output concentration C_A.

EXPERIMENTAL

A schematic illustration of the PWM setup is shown in Figure 4a. Hydrostatic pressure (0.25 psi) is evenly applied to reservoirs of different media (concentration C_A and C_B) to keep their flow rate equal and constant over time. The switching between the two input media is performed by a computer controlled 3-way solenoid valve. The valve's outlet is connected to the cell imaging chamber (μ-slide VI, ibidi) through PTFE tubing that serves as a diffusion channel to filter out individual PWM packages and create the desired media concentration. Fluorescence dye (FITC-Dextran or Alexa 680-Dextran M.W. = 3000, Invitrogen) was diluted in medium A to visualize output concentrations. Images of budding yeast cells were acquired on a fully automated inverted epi-fluorescence microscope (Ti-Eclipse, Nikon) in a temperature incubator set to 30 °C, using 60x oil objectives and appropriate excitation and emission filters. For Hog1 relocation experiments, the concentration of NaCl is steadily increased (0.05 - 0.4 M), microscopic images are taken in multiple positions. For image analysis, individual cells were tracked using segmentation of the Hta2-CFP images to identify their nuclei, while the cell area was obtained by segmentation of the RFP image or defocused transmission image.

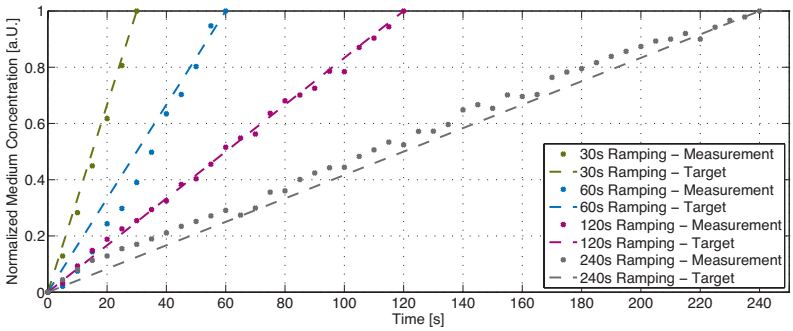

Figure 5: Experimental validation of concentration ramping using fluorescence microscopy. Fluorescent dye was diluted in input medium A to visualize the output concentration. Thus, the average pixel intensity value of the resulting images represents the medium concentration diluted by PWM. Intensity values are normalized for comparison.

RESULTS AND DISCUSSION

At first, we assessed static dilution of desired medium concentrations, as described in equation (2) with fluorescence microscopy. On the right part of Figure 4b, images of static flows of three different output concentrations of fluorescent FITC dye are presented, which are diluted as schematically illustrated in the left part of Figure 4b. Due to an increased duty cycle D, the average intensity in the microscopy images increases respectively. Next, we applied our extended version of PWM (Eq. (3)) to generate temporal concentration profiles. Throughout the ramping phase, the duty cycle D is constantly increased, leading to increased durations of the high-concentration media pulses. We performed several runs at various ramping speeds and confirmed that we can successfully ramp media concentrations using PWM (Fig. 5). Finally, we applied the PWM method to modulate the Hog1 activation by steadily increasing salt concentration (during a time period of 30 min), and observed the accumulation of Hog1 at nuclei (Fig. 6) [176].

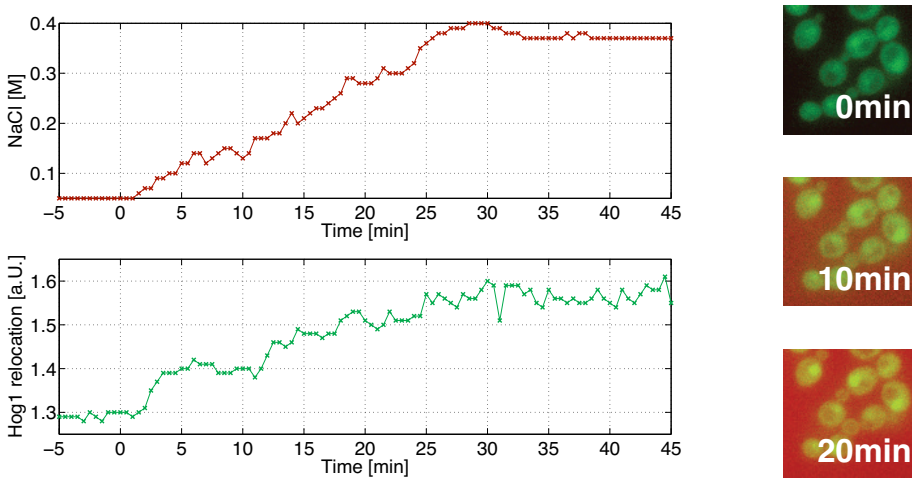

Figure 6: Hog1 nuclear relocation upon increase of salt concentration (NaCl 0.05 - 0.4 M) over duration of 30 min. *(left upper panel)* Temporal profile of NaCl concentration *(left lower panel)* Hog relocation (ratiometric average intensity ($I_{avg,\,nucleus}/I_{avg\,cyto}$) of Hog1-YFP in single cell) *(right panel)* Example images at given time points; green: Hog1-YFP, red: Alexa 680 proportional to NaCl concentration.

CONCLUSION

We introduced the concept of PWM with a dynamic duty cycle over time for liquid flows. Thus, presenting a robust and reliable way of generating temporal concentration profiles like concentration ramping. As no custom designed chips are required, it is a method easy to setup that can be integrated in many existing microfluidic environments.

ACKNOWLEDGEMENTS

This work was supported by an Interdisciplinary Pilot Project (IPP) within the Swiss Initiative in Systems Biology (SystemsX.ch). The authors would like to thank S. Pelet and H. Sharifian for donating the yeast strain.

4

MANUSCRIPT 2

REFERENCE

P. Nandy, M. Unger, C. Zechner, and H. Koeppl, *Optimal Perturbations for the Identification of Stochastic Reaction Dynamics*, in 16th IFAC Symposium on System Identification, 2012, pp. 686-91.

AUTHOR CONTRIBUTIONS

PN, MU, CZ and HK participated in designing the project. PN and MU implemented the methods and performed simulations. PN, MU, CZ and HK wrote the paper.

OPTIMAL PERTURBATIONS FOR THE IDENTIFICATION OF STOCHASTIC REACTION DYNAMICS

P. Nandy[1], M. Unger[1] C. Zechner[1], and H. Koeppl[1]

ABSTRACT

Identification of stochastic reaction dynamics inside the cell is hampered by the low-dimensional readouts available with today's measurement technologies. Moreover, such processes are poorly excited by standard experimental protocols, making identification even more ill-posed. Recent technological advances provide means to design and apply complex extracellular stimuli. Based on an information-theoretic setting we present novel Monte Carlo sampling techniques to determine optimal temporal excitation profiles for such stochastic processes. We give a new result for the controlled birth-death process and provide a proof of principle by considering a simple model of regulated gene expression.

KEYWORDS

Optimal Experiment Design, Excitation, Identifiability, Parameter Estimation, Identification Algorithms

Introduction

Advances in experimental techniques of molecular biology enable new ways to manipulate and control cells. Such methods are leverages of the emerging field of synthetic biology – the forward engineering of molecular biology from small standardized parts. For instance, with the introduction of the light-inducible promoter based on the LOV protein-domain [47], a particular gene regulatory program can conveniently be switched on by exposing cells to light of a specific spectrum (see [252, 206]). Other efforts involve the rewiring of signal cascades to induce gene expression by new or alternative extra-cellular stimuli [87]. In this light, the control of single cell dynamics comes within reach [157, 230]. In turn, these techniques also provide new ways to excite intra-cellular biochemical networks in a possibly persistent sense. The networks' precise wiring is still to a large extent unkown and their reverse-engineering is hampered by non-identifiability issues that partially are due to the lack of persistent excitation. This paper is dedicated to that issue. In particular, we address optimal experiment design for stochastic chemical kinetics.

1 Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, 8092 Zurich, Switzerland (e-mail: koeppl@ethz.ch)

Experiment design has a long-standing tradition in chemical process engineering to optimize the chemistry within batch-reactors [70]. Recently, several authors have started to address the problem of input stimuli design or optimal experimental design for intra-cellular dynamics. For instance, [18] gave a remarkable proof-of-principle what is possible with well-designed input stimuli in the setting of parameter estimation. They were able to reduce the variance of their estimates 60-fold. Single-molecule counting techniques, such as Fluorescence *In Situ* Hybridization (FISH) [64], together with the mentioned methods to directly control the promoter activity by an exogenous input (e.g. light) provides a well-defined setup to resolve stochastic transcriptional dynamics and gene-regulatory mechanisms in general. To the best of the authors knowledge no prior work in this field exists.

Classically, the theory of experimental design is centered around the Fisher Information Matrix (FIM), that can roughly be thought of as the sensitivity of the likelihood (or cost) function with respect to the model parameters [63]. A low sensitivity means that the corresponding parameter is weakly determined by the experimental data that entered the likelihood function. This generally translates into high variances if one wants to estimate this parameter from data. The covariance of any unbiased estimator is bounded from below by the inverse of the FIM. This classical approach has been applied in systems biology research (see [123]), however it has certain limitations. An important one with respect to systems biology is that the framework assumes model matching in the sense that the FIM needs to be evaluated at the true parameter value. The novel robust design approach of [66] addresses part of the problem by accounting for the large uncertainty in estimates of kinetic constants. Another, more subtle drawback is that it is a local approach and is based on second order statistics. Indeed, the FIM appears in the second term of a Taylor series expansion of the Kullback-Leibler (KL) divergence between two conditional densities (or likelihoods) evaluated at two close parameter sets [125]. Information theoretic approaches based directly on the KL divergence [138, 215] are more general. The divergence provides a measure of likelihood "tightness" that does not rely on order statistics. However, such approaches are often computationally demanding. In particular, one may need to apply Markov Chain Monte Carlo (MCMC) techniques to sample the respective densities [121], if one wants to refrain from approximations. A Bayesian counterpart to the frequentist FIM approach is also proposed [39]. Here we outline an information-theoretic approach dedicated to the identification of stochastic kinetics where the state is described by a Continuous Time Markov Chain (CTMC).

The remaining part of the work is organized as follows. Section *Optimal input based on complete observations* starts with introducing basic properties of a CTMC that are used throughout the work before the chosen experiment design framework is discussed and the result for full-path observations is presented. In Section *Optimal input sequence for the birth-death process*, we

present a proof that an optimal input sequence for a Birth/Death model can be found within the assumption of complete observations. Section *The Case of Noisy and Incomplete Measurements* generalizes the setting to the realistic scenario of having sub-sampled and noisy observations. Simulation studies, that should serve as a proof-of-principle, are discussed in Section *Simulation Results* for small stochastic models. We finally draw conclusion in Section *Discussion*.

Optimal input based on complete observations

We are in the setting of noise-free observations of full sample paths from a CTMC $\mathbf{X}$ in the time interval $[0, T]$. Let $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))'$ be the vector whose i-th element represents the number of type i species in the system at time t. Since the state-space of $\mathbf{X}$ is $\mathbb{Z}_+^n$, a sample path of $\mathbf{X}$ is determined by its jump times and the types of jumps at those times. We consider a n-species reaction system which consists of ν reaction channels with rates $c_1, c_2, \dots, c_\nu$ and corresponding reaction hazards $h_1(\mathbf{z}, c_1), \dots, h_\nu(\mathbf{z}, c_\nu)$ where $\mathbf{z} \in \mathbb{Z}_+^n$. We define $\mathbf{c} = (c_1, c_2, \dots, c_\nu)'$ and the combined hazard $h(\mathbf{z}, \mathbf{c}) = \sum_{j=1}^{\nu} h_j(\mathbf{z}, c_j)$.

Let $\tau_1, \tau_2, \dots$ be the jump times of the Markov chain $\mathbf{X}$, that is assuming $\tau_0 = 0$, we define inductively

$$\tau_k = \inf\{s > \tau_{k-1} \mid \mathbf{X}(s) \neq \mathbf{X}(\tau_{k-1})\}.$$

Note that a jump can happen at τ_k if and only if a reaction occurs at that time and

$$\tau_k - \tau_{k-1} \mid (\mathbf{X}(\tau_{k-1}) = \mathbf{x}_{k-1}) \sim \mathrm{Exp}(h(\mathbf{x}_{k-1}, \mathbf{c})).$$

If γ_k denotes the reaction at τ_k, then

$$P(\gamma_k = j \mid \mathbf{X}(\tau_{k-1}) = \mathbf{x}_{k-1}) = \frac{h_j(\mathbf{x}_{k-1}, c_j)}{h(\mathbf{x}_{k-1}, \mathbf{c})}$$

and γ_k and τ_k are conditionally independent given $\mathbf{X}(\tau_{k-1})$.

Let $\mathbf{x}$ be a realization of the Markov chain $\mathbf{X}$ in the time interval $[0, T]$. It can be fully characterized by the number of reactions which occurred in the time interval $[0, T]$ (we denote the number by M) and time and type of each reaction event, $(\tilde{\tau}_i, \tilde{\gamma}_i)$, $i = 1, 2, \dots M$, where the $\tilde{\tau}_i$ are assumed to be in increasing order and $\tilde{\gamma}_i \in \{1, 2, \dots \nu\}$. We also define $\tilde{\tau}_0 = 0$ and $\tilde{\tau}_{M+1} = T$. We will use the notation $p(Z)$ to denote the density of a random variable Z unless stated otherwise. Then the complete-data likelihood for the described stochastic kinetic model on the time interval $[0, T]$ takes the form

$$p(\mathbf{x} \mid \mathbf{c}) = \prod_{i=1}^{M} h_{\tilde{\gamma}_i}(\mathbf{x}(\tilde{\tau}_{i-1}), c_{\tilde{\gamma}_i}) \times \exp\left\{ -\sum_{i=0}^{M} h(\mathbf{x}(\tilde{\tau}_i), \mathbf{c})[\tilde{\tau}_{i+1} - \tilde{\tau}_i] \right\}. \quad (4)$$

In the case of simple mass-action kinetic rate laws the hazard function can be written in the form $h_j(\mathbf{z}, c_j) = c_j g_j(\mathbf{z})$. Let r_j denote the number of reaction events of type j that occurred in the sample path $\mathbf{x}$. Assuming a conjugate Gamma prior [245] over the kinetic parameters $\mathbf{c} = (c_1, c_2, \ldots, c_\nu)'$ ($c_j \sim \Gamma(a_j, b_j)$ and c_j's are independently distributed), the posterior density of $\mathbf{c}$ given a full path observation becomes

$$p(\mathbf{c} \mid \mathbf{x}) = \prod_{j=1}^{\nu} p(c_j \mid \mathbf{x}),$$

where $p(c_j \mid \mathbf{x})$ is the density of $\Gamma(a_j^*, b_j^*)$ with $a_j^* = a_j + r_j$ and $b_j^* = b_j + \int_0^T g_j(\mathbf{x}(t))dt = b_j + \sum_{i=0}^{M} g_j(\mathbf{x}(\tilde{\tau}_i))[\tilde{\tau}_{i+1} - \tilde{\tau}_i]$.

In an experiment, new paths can be observed but also controlled. Those paths are parametrized by an input disturbance u, out of a set of disturbances $\mathcal{U} \subseteq D[0, T]$, where D is the set of piecewise constant positive functions with finite number of jumps. We further assume that the following conditions hold.
Bounded input:

$$0 \leqslant u(t) \leqslant \bar{U} \qquad \forall t \in [0, T] \text{ and } \forall u \in \mathcal{U}. \tag{5}$$

Constant energy level of the input signal (L^1-constraint):

$$\int_0^T u(t)dt = I \qquad \forall u \in \mathcal{U}. \tag{6}$$

Most of the results and algorithms presented can be extended to a different class of admissible inputs (e.g. continuous functions). Let $\mathbf{x}_u$ denote a realization of the CTMC $\mathbf{X}_u$ when input u is applied to the system. Among all $u \in \mathcal{U}$ we want to pick those to generate the next observation paths that maximize the expectation (over all paths) of the KL divergence

$$D_{KL}\left[p(\mathbf{c} \mid \mathbf{x}_u^{(2)}, u) \,\|\, p(\mathbf{c} \mid \mathbf{x}^{(1)})\right].$$

Let $u(t)$ denote the value of the input at time t. For the next experiment assume that the value of a particular rate constant (say the k-th rate constant) changes over time by the following equation

$$c_k(t) = c_k u(t). \tag{7}$$

The complete-data likelihood for this model takes the form

$$p(\mathbf{x}_u \mid \mathbf{c}, u) = \prod_{i=1}^{M} c_{\tilde{\gamma}_i} \alpha_{\tilde{\gamma}_i}(\tilde{\tau}_{i-1}) g_{\tilde{\gamma}_i}(\mathbf{x}_u(\tilde{\tau}_{i-1})) \times \exp\left\{-\int_0^T \left(\sum_{j=1}^{\nu} c_j \alpha_j(t) g_j(\mathbf{x}_u(t))\right) dt\right\}, \tag{8}$$

where $\alpha_j(t) = u(t)1_{\{j=k\}} + 1_{\{j\neq k\}}$. Note that equation (8) holds for any integrable input function u and for $u \in \mathcal{U}$, the integral in the expression is just a summation since u and $\mathbf{x}_u$ are piecewise constant functions. We define $\pi(\mathbf{c}) := p(\mathbf{c} \mid \mathbf{x}^{(1)})$ to be the prior density.

Therefore, we have,

$$p(\mathbf{c} \mid \mathbf{x}_u^{(2)}, u) = \prod_{j=1}^{m} p(c_j \mid \mathbf{x}_u^{(2)}, u),$$

with $c_j \mid (\mathbf{x}_u^{(2)}, u)$

$$\sim \Gamma\left(a_j^{(1)} + r_j^{(2)},\ b_j^{(1)} + \int_0^T g_j(\mathbf{x}_u^{(2)}(t))\alpha_j(t)dt\right),$$

where $c_j \mid \mathbf{x}^{(1)} \sim \Gamma(a_j^{(1)}, b_j^{(1)})$.

Let $\mathbf{c}^*$ be the vector of unknown rate constants. Then our objective function should be

$$J^*(u, \mathbf{c}^*) = \int_{\mathcal{X}} \left[\int_{\Omega} p(\mathbf{c} \mid \mathbf{x}_u, u) \log \frac{p(\mathbf{c} \mid \mathbf{x}_u, u)}{\pi(\mathbf{c})} d\mathbf{c}\right] \times p(\mathbf{x}_u \mid \mathbf{c}^*, u)\ d\mathbf{x}_u, \tag{9}$$

where we removed clutter by defining $\mathbf{x}_u$ to be the previous $\mathbf{x}_u^{(2)}$. The integral over the CTMC paths can be viewed as integrations over the time and type of reaction events $(\tilde{\tau}_i, \tilde{\gamma}_i)$, $i = 1, 2, \ldots, M$ and then a sum over all possible choices of $M \in \{0, 1, \ldots\}$. But since it depends on unknown rate constants, the optimization problem is not feasible. Thus we move to a reasonable approximation of the objective function which does not depend on unknown rate constants. We replace $p(\mathbf{x}_u \mid \mathbf{c}^*, u)$ by $p(\mathbf{x}_u) := \int_{\Omega} p(\mathbf{x}_u \mid \mathbf{c}, u)\pi(\mathbf{c})d\mathbf{c}$ which is fairly good since we are at the second stage and hence the prior density is the posterior density of the first experiment. Therefore, the objective function becomes

$$J(u) = \int_{\Omega}\int_{\mathcal{X}} D_{KL}\left[p(\mathbf{c} \mid \mathbf{x}_u, u) \,\|\, \pi(\mathbf{c})\right] p(\mathbf{c}, \mathbf{x}_u \mid u)\ d\mathbf{x}_u\ d\mathbf{c}. \tag{10}$$

Hence, the optimal input is given by

$$u_{optimal} = \arg\max_{u\in\mathcal{U}} J(u).$$

Note that

$$\begin{aligned}
& D_{KL}\left[p(\mathbf{c} \mid \mathbf{x}_u, u) \,\|\, \pi(\mathbf{c})\right] \\
= & \int_{\Omega} p(\mathbf{c} \mid \mathbf{x}_u, u) \log \frac{p(\mathbf{c} \mid \mathbf{x}_u, u)}{\pi(\mathbf{c})} d\mathbf{c} \\
= & \int_{\mathbb{R}_+^\nu} \prod_{j=1}^{\nu} p(c_j \mid \mathbf{x}_u, u) \left\{ \sum_{j=1}^{\nu} \log \frac{p(c_j \mid \mathbf{x}_u, u)}{\pi(c_j)} \right\} \prod_{j=1}^{\nu} dc_j \\
= & \sum_{j=1}^{\nu} D_{KL}\left[p(c_j \mid \mathbf{x}_u, u) \,\|\, \pi(c_j)\right] \prod_{i \neq j} \int_{\mathbb{R}_+} p(c_i \mid \mathbf{x}_u, u) dc_i \\
= & \sum_{j=1}^{\nu} D_{KL}\left[p(c_j \mid \mathbf{x}_u, u) \,\|\, \pi(c_j)\right].
\end{aligned}$$

Thus, the objective function becomes

$$J(u) = E\left[\sum_{j=1}^{m} D_{KL}\left[p(c_j \mid \mathbf{x}_u, u) \,\|\, \pi(c_j)\right]\right]. \tag{11}$$

For each reaction rate c_j, the KL divergence between the posterior and the prior for given values of X_u can be computed easily as it is nothing but KL divergence between two gamma distributions. More precisely, if the prior and the posterior distribution of c_j are $\Gamma(a_j^{(1)}, b_j^{(1)})$ and $\Gamma(a_j^{(2)}, b_j^{(2)})$ respectively, the KL divergence is given by

$$\begin{aligned}
\log\left(\frac{\Gamma(a_j^{(1)})(b_j^{(2)})^{a_j^{(1)}}}{\Gamma(a_j^{(2)})(b_j^{(1)})^{a_j^{(1)}}}\right) + (a_j^{(2)} - a_j^{(1)})\psi(a_j^{(2)}) \\
+ a_j^{(2)} \frac{b_j^{(1)} - b_j^{(2)}}{b_j^{(2)}}.
\end{aligned} \tag{12}$$

Note that for any fixed $u \in \mathcal{U}$ and given set of parameters we are able to generate sample paths of the CTMC $\mathbf{X}_u$ (e.g. using the Gillespie algorithm), we can obtain a Monte-Carlo estimate of $J(u)$ using the following algorithm.

- Generate $(\mathbf{c}^{(1)}, \mathbf{x}_u^{(1)} \mid u), \ldots, (\mathbf{c}^{(S)}, \mathbf{x}_u^{(S)} \mid u)$ by generating $\mathbf{c}^{(r)}$ from $\pi(\mathbf{c})$ and $\mathbf{x}_u^{(r)}$ from $p(\mathbf{x}_u \mid \mathbf{c}, u)$, for $r = 1, 2, \ldots, S$.
- For each $r = 1, \ldots, S$ compute

$$\sum_{j=1}^{m} D_{KL}\left[p(c_j \mid \mathbf{x}_u^{(r)}, u) \,\|\, \pi(c_j)\right].$$

- Finally compute

$$\hat{J}(u) = \frac{1}{S} \sum_{r=1}^{S} \sum_{j=1}^{m} D_{KL}\left[p(c_j \mid \mathbf{x}_u^{(r)}, u) \,\|\, \pi(c_j)\right].$$

Now standard optimization techniques could be used to locate the optimal u (i.e., which maximizes expected the KL divergence).In case some of the rate constants are known, the summation index of equation (11) can be modified accordingly to sum only over the unknown rate constants $\mathbf{c}^* \subseteq \mathbf{c}$.

Optimal input sequence for the birth-death process

Since the objective function $J(u)$ can not be written as an explicit function of u, the analytical solution to the optimization problem under consideration is not straightforward even for fairly simple models. However, for the birth-death model we will obtain an analytical solution to the optimization problem under some minor assumptions.

Observe that maximizing the expected KL divergence is equivalent to minimizing the expected entropy of the posterior since

$$\begin{aligned} E\left[E\left(\log\frac{p(\mathbf{c}_j \mid \mathbf{x}_u)}{\pi(\mathbf{c}_j)} \mid \mathbf{x}_u\right)\right] = E[E\{\log(p(\mathbf{c}_j \mid \mathbf{x}_u) \mid \mathbf{x}_u\}] \\ -E[\log(\pi(\mathbf{c}_j))]. \end{aligned}$$

We know that if $Z \sim \Gamma(a, b)$ it follows from the central limit theorem that Z is approximately distributed as a normal random variable with mean a/b and variance $\sigma^2 := a/b^2$ provided a is large enough. We recall that $\frac{1}{2}\log(2\pi e\sigma^2)$ is the entropy of a normally distributed random variable with variance σ^2. Then the entropy of Z can be approximated as

$$\begin{aligned} H(Z) &= (1-a)\psi(a) + \log(\Gamma(a)) + a + \frac{1}{2}\log(\frac{\sigma^2}{a}) \\ &\approx \frac{1}{2}\log(2\pi e\sigma^2) \end{aligned} \tag{13}$$

for sufficiently large values of a. In fact it can be shown that $H(Z)$ converges to $\frac{1}{2}\log(2\pi e\sigma^2)$ if we let a and b tend to infinity keeping σ^2 constant. Therefore, the entropy can be approximately viewed as a monotonic function of the variance provided the shape parameter a is large enough. In our case, the shape parameter of the posterior of the j-th rate constant $a_j^{(2)} = a_j^{(1)} + r_j$ will usually have high values if we observe the system long enough and hence minimizing the expected variance could be an equivalent objective.

We consider the following birth-death model:

$$\emptyset \xrightarrow{c_1^*} A \xrightarrow{c_2^*} \emptyset \tag{14}$$

where c_1^* and c_2^* are the unknown rate constants for birth and death respectively. Now suppose we control the birth rate with input u, i.e. in the controlled system the birth rate at time t is given by $c_1^*(t) := c_1^* u(t)$.

Let $\Gamma(a_1, b_1)$ and $\Gamma(a_2, b_2)$ be the independent priors for the birth rate c_1 and the death rate c_2 respectively. Then for the input u which satisfies equation (6) the posterior of c_1 will be $\Gamma(a_1 + r_1^{(u)}, b_1 + \int_0^T u(s)ds) =$

$\Gamma(a_1 + r_1^{(u)}, b_1 + I)$ where $r_1^{(u)}$ is the total number of births for the input u. Therefore

$$r_1^{(u)} \sim \mathcal{P}(c_1^* \int_0^T u(s)ds) = \mathcal{P}(c_1^* I). \tag{15}$$

Thus the posterior distributions of $r_1^{(u)}$ and hence the expected variances of the posteriors are identical for all inputs under consideration. Now let u_1 and u_2 be two inputs which satisfy equation (6) and

$$\int_0^t u_1(s)ds \geqslant \int_0^t u_2(s)ds \quad \forall t \in [0, T]. \tag{16}$$

We first prove the following result which will be used to show that the input u_1 is better than u_2 in the sense that the expected posterior variance of c_2 for u_1 is smaller (Theorem 1).

Lemma 1 *Let* $(Z \mid X_i, Y_i) \sim \Gamma(a + X_i, b + Y_i)$ *and let* $X_i \sim \mathcal{P}(cY_i)$. *Then if* $c \leqslant 2a/b$ *and* Y_1 *is stochastically larger than* Y_2, *then*

$$E[\mathrm{Var}(Z \mid X_1, Y_1)] \leqslant E[\mathrm{Var}(Z \mid X_2, Y_2)].$$

Proof 1 *Note that,*

$$E[\mathrm{Var}(Z \mid X_i, Y_i)] = EE[\mathrm{Var}(Z \mid X_i, Y_i) \mid Y_i] = E[f(Y_i)]$$

where

$$\begin{aligned} f(\lambda) &= E[\mathrm{Var}(Z \mid X_i, Y_i) \mid Y_i = \lambda] \\ &= \frac{a + E(X_i \mid Y_i = \lambda)}{(b+\lambda)^2} = \frac{a + c\lambda}{(b+\lambda)^2}. \end{aligned}$$

Therefore

$$f'(\lambda) = \frac{bc - 2a - c\lambda}{(b+\lambda)^3} \leqslant 0, \quad \forall \lambda \geqslant 0.$$

Hence, the result follows as Y_1 *is stochastically larger than* Y_2 *and* $f(\lambda)$ *is a non-increasing function.* ■

Let $X_u(t)$ denote the population size at time t when an input u is applied to the system and u_1 and u_2 be any two inputs which satisfy equation (6) and equation (16). We assume that the initial population size, which will be denoted by $X(0)$, to be deterministic and independent of the input applied to the system. Now if T is large enough then it can be assumed that $c_2^* \leqslant \frac{2a_2}{b_2}$ and we have the following result.

Theorem 1 $\int_0^T X_{u_1}(t)dt$ *is stochastically larger than* $\int_0^T X_{u_2}(t)dt$ *and hence the expected posterior variance of* c_2 *for* u_1 *is smaller.*

Proof 2 *Let* $N_t^{(i)}$ *denote the total number of births up to time* t *for the input* u_i. *Then* $N_T^{(i)} = r_1^{(u_i)} \sim \mathcal{P}(c_1^* I)$ *and* $N_t^{(i)}$ *is an inhomogeneous poisson process with cumulative hazard* $\Lambda_i(t) = \int_0^t c_1^* u_i(s)ds$. *Thus,* $N_t^{(i)} \stackrel{d}{=} Z(\int_0^t c_1^* u_i(s)ds)$, *where* $Z(t)$ *is an unit poisson process. Now conditioning on* $N_T^{(i)} = n$,

$$\begin{aligned}&\int_0^T X_{u_i}(t)dt\\ &= \sum_{j=1}^{X(0)+n} \left[W_j^{(i)} 1_{\{W_j^{(i)} \leqslant T - t_j^{(i)}\}} + (T - t_j^{(i)}) 1_{\{W_j^{(i)} > T - t_j^{(i)}\}}\right] \end{aligned} \tag{17}$$

where $W_j^{(i)}$*'s are i.i.d.* $\mathrm{Exp}(c_2^*)$ *random variables and* $t_j^{(i)} = 0$ *for* $j \leqslant X(0)$ *and* $\{t_{X(0)+1}^{(i)}, \dots t_{X(0)+n}^{(i)}\}$ *are unordered time points of birth events. Then conditioning on* $N_T^{(i)}$, $t_{X(0)+j}^{(i)}$*'s are independent and identically distributed with the density* $u_i(s)/I$.

Therefore, $\int_0^t u_1(s)ds \geqslant \int_0^t u_2(s)ds$ *implies* $t_{X(0)+j}^{(2)}$ *is stochastically larger than* $t_{X(0)+j}^{(1)}$ *for* $j = 1, 2, \dots n$. *Thus*

$$\begin{aligned}&W_j^{(1)} 1_{\{W_j^{(1)} \leqslant T - t_j^{(1)}\}} + (T - t_j^{(1)}) 1_{\{W_j^{(1)} > T - t_j^{(1)}\}}\\ &\underset{ST}{\geq} W_1^{(2)} 1_{\{W_j^{(2)} \leqslant T - t_j^{(2)}\}} + (T - t_j^{(2)}) 1_{\{W_j^{(2)} > T - t_j^{(2)}\}}\end{aligned}$$

for $j = X(0) + 1, \dots X(0) + n$, *as for each fix* $W_j^{(i)} = w$, $f_w(t) = w 1_{\{w \leqslant T-t\}} + (T - t) 1_{\{w > T-t\}}$ *is a non-increasing function of* t *and* $W_j^{(i)}$ *and* $t_j^{(i)}$*'s are independent [200].*

Therefore, we have,

$$\left(\int_0^T X_{u_1}(t)dt\right) \Bigg| N_T^{(1)} = n \underset{ST}{\geq} \left(\int_0^T X_{u_2}(t)dt\right) \Bigg| N_T^{(2)} = n$$

for each n *as the sum in (17) is a sum of independent random variables. Now since* $N_T^{(1)}$ *and* $N_T^{(2)}$ *are identically distributed,*

$$\int_0^T X_{u_1}(t)dt \underset{ST}{\geq} \int_0^T X_{u_2}(t)dt. \tag{18}$$

We recall the posterior of c_2 *is distributed as*

$$\Gamma(a_2 + r_2^{(u_i)}, b_2 + \int_0^T X_{u_i}(s)ds)$$

for input u_i *and*

$$r_2^{(u_i)} \sim \mathcal{P}(c_2^* \int_0^T X_{u_i}(s)ds).$$

Hence, the result follows from (18) and Lemma 1 provided $c_2^* \leqslant \frac{2a_2}{b_2}$. ■

Let u_* be the input such that

$$u_*(t) = \begin{cases} \bar{U} & \text{if } t \in [0, I/\bar{U}) \\ 0 & \text{if } t \in [I/\bar{U}, T]. \end{cases}$$

Then clearly for any other input u which satisfies (5) and (6) (i.e. $0 \leqslant u(t) \leqslant \bar{U}$, $\forall t \in [0, T]$ and $\int_0^T u(s)ds = I$), we have, $\int_0^t u_*(s)ds \geqslant \int_0^t u(s)ds$ for all $t \in [0, T]$. Therefore, it follows from Theorem 1 that u_* is an optimal input for this system.

The Case of Noisy and Incomplete Measurements

In the following we describe how our method can be modified to deal with the realistic scenario where only noisy and partial measurements can be recorded at discrete time points. Let $\mathbf{y}_u := \{\mathbf{y}_u(t_1), \ldots, \mathbf{y}_u(t_N)\}$ be a set of measurements when the input u is applied to the system. Note that conditional on the latent CTMC path $\mathbf{x}_u$, the observation $\mathbf{y}_u$ is independent of the kinetic parameters and hence, is sufficiently described by the measurement likelihood function $p(\mathbf{y}_u \mid \mathbf{x}_u)$. In most practical scenarios, the acquisition uncertainty can be well modeled as a normal or log-normal distribution [122], whereas the scaling parameters of such densities are typically unknown. However, for the sake of clarity, we assume the functional form of $p(\mathbf{y}_u \mid \mathbf{x}_u)$ to be entirely known (i.e., it does not contain any unknown parameter).

Then the posterior density of the rate constants is given by

$$\begin{aligned} p(\mathbf{c} \mid \mathbf{y}_u, u) &= \int_{\mathcal{X}} p(\mathbf{c}, \mathbf{x}_u \mid \mathbf{y}_u, u) d\mathbf{x}_u \\ &= \int_{\mathcal{X}} p(\mathbf{c} \mid \mathbf{x}_u, u) p(\mathbf{x}_u \mid \mathbf{y}_u) d\mathbf{x}_u \\ &= \int_{\mathcal{X}} p(\mathbf{c} \mid \mathbf{x}_u, u) \frac{p(\mathbf{y}_u \mid \mathbf{x}_u) p(\mathbf{x}_u \mid u)}{p(\mathbf{y}_u \mid u)} d\mathbf{x}_u \\ &= \frac{E_{(\mathbf{x}_u \mid u)}[p(\mathbf{c} \mid \mathbf{x}_u, u) p(\mathbf{y}_u \mid \mathbf{x}_u)]}{E_{(\mathbf{x}_u \mid u)}[p(\mathbf{y}_u \mid \mathbf{x}_u)]}. \end{aligned}$$

Note that $p(\mathbf{c} \mid \mathbf{y}_u, u)$ involves an expectation over the latent state $\mathbf{x}_u$ and consequently, cannot be solved analytically. However, sampling techniques, such as sequential Monte-Carlo can be used to evaluate - or simulate from $p(\mathbf{c} \mid \mathbf{y}_u, u)$ [257].

Now the expected KL divergence becomes

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{D}_{\mathrm{KL}}\left[p(\mathbf{c} \mid \mathbf{y}_u, u) \,\|\, \pi(\mathbf{c})\right]\right] \\
&= \int\int_{\Omega} p(\mathbf{y}_u \mid u)\; p(\mathbf{c} \mid \mathbf{y}_u, u)\; \log \frac{p(\mathbf{c} \mid \mathbf{y}_u, u)}{\pi(\mathbf{c})}\; d\mathbf{c}\; d\mathbf{y} \\
&= \int\int_{\Omega} p(\mathbf{y}_u, \mathbf{c} \mid u) \times \\
&\quad \log \frac{\mathrm{E}_{(\mathbf{x}_u \mid u)}[p(\mathbf{c} \mid \mathbf{x}_u, u) p(\mathbf{y}_u \mid \mathbf{x}_u)]}{\mathrm{E}_{(\mathbf{x}_u \mid u)}[p(\mathbf{y}_u \mid \mathbf{x}_u)]\pi(\mathbf{c})}\; d\mathbf{c}\; d\mathbf{y} \\
&= \mathrm{E}_{(\mathbf{y}_u, \mathbf{c} \mid u)}\left[\log \frac{\mathrm{E}_{(\mathbf{x}_u \mid u)}[p(\mathbf{c} \mid \mathbf{x}_u, u) p(\mathbf{y}_u \mid \mathbf{x}_u)]}{\mathrm{E}_{(\mathbf{x}_u \mid u)}[p(\mathbf{y}_u \mid \mathbf{x}_u)]}\right] + K
\end{aligned}$$

with K as a constant term, independent of u, such that we define

$$J(u) := \mathrm{E}_{(\mathbf{y}_u, \mathbf{c} \mid u)}\left[\log \frac{\mathrm{E}_{(\mathbf{x}_u \mid u)}[p(\mathbf{c} \mid \mathbf{x}_u, u) p(\mathbf{y}_u \mid \mathbf{x}_u)]}{\mathrm{E}_{(\mathbf{x}_u \mid u)}[p(\mathbf{y}_u \mid \mathbf{x}_u)]}\right].$$

Estimates of $J(u)$ can be computed as follows:

- Generate $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(Q)}$ from $\pi(\mathbf{c})$.
- For each $r = 1, 2, \ldots, Q$; generate $\mathbf{x}_u^{(r)}$ from $p(\mathbf{x}_u \mid \mathbf{c}^{(r)}, u)$.
- Generate $\mathbf{y}_u^{(i)}$ from $p(\mathbf{y}_u \mid \mathbf{x}_u^{(i)})$, for $i = 1, \ldots, Q$.
- Compute

$$\hat{J}(u) = \frac{1}{Q} \sum_{j=1}^{Q} \log \frac{\sum_{i=1}^{Q} p(\mathbf{c}^{(j)} \mid \mathbf{x}_u^{(i)}, u) p(\mathbf{y}_u^{(j)} \mid \mathbf{x}_u^{(i)})}{\sum_{i=1}^{Q} p(\mathbf{y}_u^{(j)} \mid \mathbf{x}_u^{(i)})}.$$

We want to point out that it is straight-forward to replace the KL divergence by other objective functions (e.g., the generalized variance) in the computation above.

Remark 1 *If we want to apply our input design technique at the second stage (or later) then the prior $\pi(\mathbf{c})$ is not Gamma anymore (as it would be the posterior of the previous experiment). As mentioned earlier, samples from the posterior are typically drawn using Monte-Carlo algorithms, such that the prior for the subsequent experiment is represented by a finite set of samples. In such cases we can still draw from $\pi(\mathbf{c})$, evaluation - however - requires suitable approximations.*

Simulation Results

We use two simulation examples to illustrate our method for the case of complete observations (see Section *Optimal input based on complete observations*). In the first example, we consider the birth-death model where the simulation result shows that the optimal input obtained in Section *Optimal input sequence for the birth-death process* leads to significant information gain

in comparison to a standard reference input. We then consider an abstract transcription model depicted in Figure 7 which involves 5 species and 7 reaction channels and apply our method to compute estimates of the expected KL divergence for a proposed input sequence and finally choose the optimum one among the proposed inputs.

To limit the space of possible input proposals u_i, we used piecewise constant sequences with N segments of time duration ΔT, each constrained as given by equations (5) and (6). New input sequences were proposed and accepted based on uniform sampling over the space of admissible inputs. We selected the best proposed sequence based on the maximal (estimated) value of the objective function defined in equation (10). Other common techniques for locating an approximate global optimum such as simulated annealing can also be used if the search space is very large. As a validation step where we compute the expected KL divergence for fixed rate constants (see equation (9)), we applied the optimized input sequence u_{opt} and a reference *Step-Up* sequence u_{ref}, a standard input stimulus in dynamic cell experiments, to the model. The results are summarized in the tables below. Note that the Kullback-Leibler divergence is based on the logarithm of a likelihood, and thus small increases thereof can be significant in terms of information gain.

Birth-Death Process:

The birth and the death rate constants were chosen to be $c_1^* = 0.5$ and $c_2^* = 0.4$ respectively. We took independent gamma priors with identical mean 1 and variances 1 and 1.25 respectively for the birth and the death rate. Then we computed Monte-Carlo estimates of the expected KL divergence and the posterior variance of the death rate c_2 for $u_{ref} = [0, 0, 0, 5, 5, 5]$ and the optimal input u_{opt} derived in section *Optimal input sequence for the birth-death process*. The estimates of the expected KL divergences and the expected posterior variances which are given in Table 1, are based on $20,000$ independent simulations with $N = 6$, $\Delta T = 1$ and $\bar{U} = 5$. We also obtained an estimate of the standard error in each case using bootstrapping.

Input sequence	$\hat{J}^*(u, \mathbf{c}^*)$ (in 10^{-2})	Expected posterior variance (in 10^{-2})
$[0, 0, 0, 5, 5, 5]$	106.80 (0.09)	3.93 (0.019)
$[5, 5, 5, 0, 0, 0]$	121.47 (0.11)	2.65 (0.013)

Table 1: Estimates of the expected KL divergences and the expected posterior variances of the death rate of the birth-death process for u_{ref} and u_{opt}. The numbers in the brackets are the associated standard deviations.

Transcription Model:

Input sequences ($m = 500$) for a simple model of regulated gene expression (Figure 7) were uniformly drawn from the proposal space with $N = 6$,

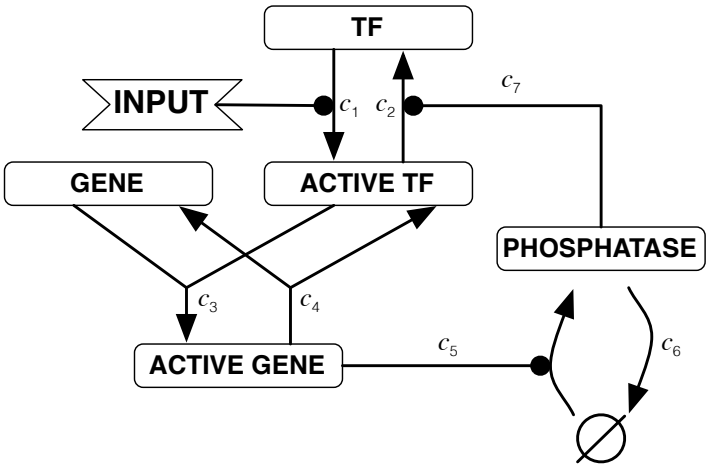

Figure 7: An abstract model of regulated gene expression. The input changes the rate of transcription factor (TF) activation. An active TF can bind to a gene, resulting in an increase of phosphatase copies, which close an inhibitory feedback loop by resetting active TFs to their non-active state.

$\Delta T = 5$ and $\bar{U} = 5$ and only integer values allowed.
$\mathbf{c}^* = [0.02, 0.005, 0.2,\ 1, 0.9, 0.01, 0.005]$ was chosen to be the true value of the parameter vector. Independent gamma priors with mean c_j^* and variance $2c_j^*$ for the rate c_j were chosen. We obtained that $u_{opt} = [4, 2, 3, 2, 0, 4]$ maximizes the objective function $J(u)$ among the proposed input sequences. We summarize the results in Table 2. It can be clearly seen that on average significantly more informative data can be obtained from an experiment by applying u_{opt} in comparison to the standard perturbation.

Input sequence	$\hat{J}(u)$	$\hat{J}^*(u, \mathbf{c}^*)$
$[0, 0, 0, 5, 5, 5]$	4.0610 (0.0161)	17.8065 (0.0499)
$[4, 2, 3, 2, 0, 4]$	4.3937 (0.0182)	23.3871 (0.0362)

Table 2: Estimates of the values of the objective function and the expected KL divergences of the rate parameter vector involved in the transcription model (Figure 7) for u_{ref} and u_{opt}. The numbers in the brackets are the associated standard deviations.

Discussion

We introduced an experiment design technique for input stimuli in biochemical experiments, allowing us to propose and select temporal perturbation sequences based on information-theoretic measures. The framework was derived for the simplistic case of complete and noise-free observations, as well as for the realistic experimental scenario of noisy and discrete-time measurements. For the former case, we derived analytical

solutions for a simple birth-death process and simulation studies were additionally performed for a more complicated gene expression model which shows that significant information gain from experiments can be achieved by applying the optimal perturbation to the system. While the provided techniques were built upon a powerful mathematical framework, they are computationally very demanding. Consequently, efficient implementations and sampling techniques will be of vital importance to cover a large input proposal space and to be able to deal with real world models.

5

MANUSCRIPT 3

REFERENCE

C. Zechner, P. Nandy, M. Unger, and H. Koeppl, *Optimal variational perturbations for the inference of stochastic reaction dynamics*, in 51st IEEE Conference on Decision and Control, 2012, pp. 5336-41.

AUTHOR CONTRIBUTIONS

CZ, PN, MU and HK participated in designing the project. CZ implemented the methods and performed simulations. CZ, MU and HK wrote the paper.

OPTIMAL VARIATIONAL PERTURBATIONS FOR THE INFERENCE OF STOCHASTIC REACTION DYNAMICS

C. Zechner[1], P. Nandy[1] M. Unger[1], and H. Koeppl[1]

ABSTRACT

Although single-cell techniques are advancing rapidly, quantitative assessment of kinetic parameters is still characterized by ill-posedness and a large degree of uncertainty. In many standard experiments, where transcriptional activation is recorded upon application of a step-like external perturbation, cells almost instantaneously adapt such that only a few informative measurements can be obtained. Consequently, the information gain between subsequent experiments or time points is comparably low, which is reflected in a hardly decreasing parameter uncertainty. However, novel microfluidic techniques can be applied to synthesize more sophisticated perturbations to increase the informativeness of such time-course experiments. Here we introduce a mathematical framework to design optimal perturbations for the inference of stochastic reaction dynamics. Based on Bayesian statistics, we formulate a variational problem to find optimal temporal perturbations and solve it using a stochastic approximation algorithm. Simulations are provided for the realistic scenario of noisy and discrete-time measurements using two simple reaction networks.

Introduction

Transcriptional and post-transcriptional processes exhibit significant stochasticity, attributed partly to the molecular noise caused by low-copy molecules [61]. Classical modeling approaches based on the reaction-rate equation cannot properly capture the dynamics of such processes and a stochastic description is in order - such as provided by the Continuous Time Markov Chain (CTMC) framework. While computational challenges arise in such cases, single-cell measurements, revealing the molecular stochasticity were shown to provide a rich source of information in the context of parameter identification [160]. Experimental single cell techniques such as fluorescence microscopy or Fluorescence *In Situ* Hybridization (FISH) [64] allow quantification of transcriptional and post-transcriptional processes and their stochastic variability. Although such techniques are advancing rapidly, they can only account for a few readouts during a single experiment. Consequently, identification of the usually high-dimensional process remains highly ill-posed in most scenarios.

However, in combination with novel microfluidic techniques to generate input stimuli [253, 11, 232], one can still produce informative data from these processes. More specifically, in an earlier study [232] we developed

1 Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, 8092 Zurich, Switzerland {zechner,nandy,unger,koeppl}@control.ee.ethz.ch

methods allowing for the synthesis of rapidly changing cellular microenvironment – opening up ways to chemically induce intra-cellular processes in a complex time-varying manner.

Traditional experimental design techniques rest upon a sensitivity analysis of the parameter likelihood, represented by the Fisher Information Matrix (FIM) [63]. Experimental optimality is specified by means of a certain objective - or *utility* - function, which is extracted from the FIM. Most common choices of such functions are the trace or determinant, often associated with the nomenclatures T- and D-optimality, respectively. First promising applications to biochemical network identification are provided in [123] or [18]. In particular the results in [18], convincingly demonstrate the usefulness of optimal experimental design approaches. Extension to a Bayesian treatment seems natural, as it allows to incorporate prior knowledge from a foregoing experiment. In this case the parameter likelihood function is replaced by its Bayesian analogon, i.e., the posterior distribution (see [39] for an overview).

The authors of [120] address the problem of sensitivity- and identifiability analysis for the case of stochastic reaction dynamics, whereas - in this context - the problem of optimal perturbation design remains unsolved. A preliminary theoretical study of the use of Bayesian optimal experimental design for the inference of stochastic reaction networks is given in [162]. Therein, a purely Monte-Carlo sampling approach is used to determine the optimal perturbation if one is given complete observations of the CTMC sample paths. The high-dimensionality of the sampling problem and the assumption of complete observations limits the method's scalability and practical relevance, respectively. Here we extend this framework and formulate the problem as a variational optimization problem and deploy stochastic approximation methods for its solution. We formulate and solve the complete observation case, but importantly, also the incomplete and noisy observation case.

The remaining part of the paper is structured as follows. In Section *Stochastic Reaction Dynamics* we briefly introduce the CTMC description of stochastic chemical kinetics and discuss Bayesian parameter inference for the cases of complete - as well as incomplete observations, before the experimental design and optimization framework is introduced in Section *Variational Perturbation Design*. Simulation results are provided in Section *Simulations* for the realistic scenario of noisy measurements at discrete time points.

Stochastic Reaction Dynamics

Consider a CTMC $\mathbf{X}$ on the time interval $[0, T]$. The instances of species $i \in \{1, \dots, n\}$ at time t are represented in the i-th element of the vector $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))^T$ and change according to ν reaction channels with rates $c_1, c_2, \dots, c_\nu$ and corresponding reaction hazards $h_1(\mathbf{z}, c_1), \dots, h_\nu(\mathbf{z}, c_\nu)$, where $\mathbf{z} \in \mathbb{Z}_+^n$. We define $\mathbf{c} = (c_1, c_2, \dots, c_\nu)^T$ and the combined haz-

ard $h(\mathbf{z}, \mathbf{c}) = \sum_{j=1}^{\nu} h_j(\mathbf{z}, c_j)$ and denote $\tau_1, \tau_2, \ldots$ the jump times of the Markov chain $\mathbf{X}$. Assuming $\tau_0 = 0$, we inductively define

$$\tau_k = \inf\{s > \tau_{k-1} \mid \mathbf{X}(s) \neq \mathbf{X}(\tau_{k-1})\}.$$

A jump can happen at τ_k if and only if a reaction occurs at that time and

$$\tau_k - \tau_{k-1} \mid \mathbf{X}(\tau_{k-1}) = \mathbf{x}_{k-1} \ \sim \ \mathrm{Exp}(h(\mathbf{x}_{k-1}, \mathbf{c})).$$

If γ_k denotes the reaction at τ_k, then

$$P(\gamma_k = j \mid \mathbf{X}(\tau_{k-1}) = \mathbf{x}_{k-1}) = \frac{h_j(\mathbf{x}_{k-1}, c_j)}{h(\mathbf{x}_{k-1}, \mathbf{c})}$$

and γ_k and τ_k are conditionally independent given $\mathbf{X}(\tau_{k-1})$.

Let $\mathbf{x}$ be a realization of the Markov chain $\mathbf{X}$ in the time interval $[0, T]$. Since the state-space of $\mathbf{X}$ is $\mathbb{Z}_+^n$, the sample path $\mathbf{x}$ can be fully characterized by the initial state $\mathbf{x}_0$, the sequence of reactions occurred in the time interval $[0, T]$ (we will denote the number of occurrences by M) and the time and type of each reaction event $(\tilde{\tau}_i, \tilde{\gamma}_i)$, $i = 1, 2, \ldots, M$. We assume an increasing order of $\tilde{\tau}_i$ between $\tilde{\tau}_0 = 0$ and $\tilde{\tau}_{M+1} = T$ and $\tilde{\gamma}_i \in \{1, 2, \ldots, \nu\}$. Unless explicitly stated, the notation $p(Z)$ will be used to denote the density of a random variable Z.

Complete Observations

First, we consider noise-free observations of full sample paths. Then, the complete-data likelihood for the described stochastic kinetic model on the time interval $[0, T]$ takes the form

$$p(\mathbf{x} \mid \mathbf{c}) = \left\{ \prod_{i=1}^{M} h_{\tilde{\gamma}_i}(\mathbf{x}(\tilde{\tau}_{i-1}), c_{\tilde{\gamma}_i}) \right\} \times \exp\left\{ -\sum_{i=0}^{M} h(\mathbf{x}(\tilde{\tau}_i), \mathbf{c})[\tilde{\tau}_{i+1} - \tilde{\tau}_i] \right\} \quad (19)$$

[245]. For mass-action kinetics, the hazard function can be written as $h_j(\mathbf{z}, c_j) = c_j g_j(\mathbf{z})$. The number of type j reaction events which occurred in a sample path $\mathbf{x}$ is denoted as r_j. Using a conjugate Gamma prior over the kinetic parameters $\mathbf{c}$ (assuming $c_j \sim \Gamma(a_j, b_j)$ and c_j's are independently distributed), the posterior density of $\mathbf{c}$ given a full path observation becomes

$$p(\mathbf{c} \mid \mathbf{x}) = \prod_{j=1}^{\nu} p\left(c_j \mid \mathbf{x}\right)$$

[245], where $p(c_j \mid \mathbf{x})$ is the posterior density of $\Gamma(a_j^*, b_j^*)$ with

$$a_j^* = a_j + r_j$$

and

$$b_j^* = b_j + \int_0^T g_j(\mathbf{x}(t))dt = b_j + \sum_{i=0}^{M} g_j(\mathbf{x}(\tilde{\tau}_i))[\tilde{\tau}_{i+1} - \tilde{\tau}_i] \ .$$

Incomplete Observations

Within realistic experimental conditions, only a d-dimensional ($d \leqslant n$) and noisy readout of the CTMC $\mathbf{X}$ can be obtained at discrete time points, i.e., $\mathbf{X}$ is said to be *latent*. Let $\mathbf{y}_l \in \mathbb{R}^d$ be the measurement at time t_l for $l = 1, \ldots, L$. We assume existence of an arbitrary but known measurement likelihood function $p(\mathbf{y}_l|\mathbf{x}(t_l))$, characterized by a measurement equation, e.g., such as

$$\mathbf{y}_l = W\mathbf{x}(t_l) + \epsilon_l, \tag{20}$$

with $W \in \mathbb{R}^{d\times n}$ and i.i.d. acquisition noise ϵ_l. In this case, the joint density function over all quantities is given by

$$\begin{aligned} &p\left(\mathbf{c}, \mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_L\right) = \\ &\qquad \prod_{l=1}^{L} p\left(\mathbf{y}_l \mid \mathbf{x}(t_l)\right) p\left(\mathbf{x} \mid \mathbf{c}\right) p\left(\mathbf{c}\right) \end{aligned} \tag{21}$$

and the posterior distribution over the parameters is given by

$$\begin{aligned} &p\left(\mathbf{c} \mid \mathbf{y}_1, \ldots, \mathbf{y}_L\right) \propto \\ &\qquad \int p\left(\mathbf{c}, \mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_L\right) d\mathbf{x}, \end{aligned} \tag{22}$$

where the integral over the CTMC paths can be viewed as integrations over the time and type of reaction events $(\tilde{\tau}_i, \tilde{\gamma}_i)$, $i = 1, 2, \ldots, M$ and then a sum over all possible choices of $M \in \{0, 1, \ldots\}$. In contrast to the case of complete observations, the posterior distribution (22) cannot be computed analytically due to its complicated structure. Significant efforts have been made in literature to efficiently sample from (22) [247, 84, 8, 257]. In this work, we follow a Sequential Monte Carlo (SMC) approach, such as proposed in [257].

Variational Perturbation Design

For the moment let us consider a generic observation $\mathbf{z}$ that corresponds to either $\mathbf{x}$ in case of complete - or $\mathbf{y}$ in case of incomplete observations.

Furthermore we assume that the target system can be perturbed on the acquisition interval $[0, T]$, such that the sample paths can be written as a function of the perturbation $u \in \mathcal{U}$, i.e., $\mathbf{x}_u = \{\mathbf{x}(t, u) \mid t \in [0, T]\}$. For instance, we could assume that the perturbation allows to modulate a reaction rate such that $c_k(t) = c_k u(t)$, where $u(t)$ denotes the value of the perturbation at time t. Among all u we want to choose the one that generates maximally *informative* measurements.

Choosing the Objective Function

In the following, we assume a sequence of consecutive experiments $j = 1, 2, \ldots$ whereas the inferred posterior distributions for experiment j are

used as prior distributions for experiment $j+1$, giving rise to a recursive Bayesian experimental design scheme.

Based on information theoretic considerations, the information gain of a respective experiment $j+1$ with measurement $\mathbf{z}$ can be quantified by the negative expectation of the Kullback-Leibler (KL) divergence

$$J(u) = -E_{\mathbf{z}}\left[D_{KL}\left[p(\mathbf{c} \mid \mathbf{z}, u) \parallel \pi(\mathbf{c})\right]\right], \tag{23}$$

where the expectation is calculated over $\mathbf{z}$. The density of $\mathbf{z}$ is $p(\mathbf{z}) := \int p(\mathbf{z} \mid \mathbf{c}, u)\pi(\mathbf{c})d\mathbf{c}$ and the prior $\pi(\mathbf{c})$ is given by the posterior of the previous experiment j.

Although it allows for an elegant information theoretic interpretation, the KL divergence is often difficult to handle because of its intricate analytical form - such as in case of stochastic chemical kinetics. A more traditional objective function for experimental design purposes is the expected logarithm of the generalized posterior variance [39], i.e. the expected logarithm of the determinant of the variance-covariance matrix

$$J(u) = E_{\mathbf{z}}\left[\log|\Sigma|\right] \text{ with } \Sigma = E_{\mathbf{c}|\mathbf{z},u}\left[\mathbf{c}\mathbf{c}^T\right] - \mu\mu^T, \tag{24}$$

where $E_{\mathbf{c}|\mathbf{z},u}$ denotes a conditional expectation with respect to the posterior distribution $p(\mathbf{c} \mid \mathbf{z}, u)$ and $\mu = E_{\mathbf{c}|\mathbf{z},u}[\mathbf{c}]$. Recall that for a multivariate normal distribution the entropy is the logarithm of its generalized variance. Consequently, it can be shown that the minimizer u^* of (24) converges to the minimizer of (23), if $p(\mathbf{c} \mid \mathbf{z}, u)$ approaches a Gaussian distribution. Throughout this work, we chose equation (24) as the objective function.

Note that even though (24) exhibits simpler expressions than (23), it involves complicated expectations and has to be evaluated using Monte Carlo simulation as outlined in the following. In the complete observation case, this can be accomplished by drawing a set of rate constants $\mathbf{c}^{(i)}$ from the prior distribution, which is then used to simulate a sample path $\mathbf{x}_u^{(i)}$ (given some perturbation u). The posterior distribution (which is a Gamma distribution) and the variance-covariance matrix $\Sigma^{(i)}$ are extracted from the sample path. This procedure is repeated K times in order to estimate (24) as $1/K \sum_{i=1}^{K} \log\left|\Sigma^{(i)}\right|$. In a similar manner, the expectation can be computed in case of noisy and incomplete measurements, where additional discretization and simulation steps are performed to sample $\mathbf{y}_l^{(i)}$ conditionally on $\mathbf{x}_u^{(i)}(t_l)$. Note - however - that this requires more Monte Carlo runs to properly account for the additional sampling dimensions. Furthermore - as indicated in (22) - each run involves an integration over the path space yielding a significant increase of computational effort. As a possible solution, certain simplifications can be made to reduce the problem complexity. For instance, the expectation over the sample paths and measurements could be "moved into the computation" of $\Sigma^{(i)}$, yielding the generalized log-variance of the expected observation. We want to point out that in this case, the resulting perturbation design cannot not account for any process and acquisition variability and consequently, does

not provide a viable alternative. During our simulation studies from Section *Simulations* we observed that an efficient and accurate strategy is to move only the expectations over ϵ_l into the calculation of $\Sigma^{(i)}$, while leaving the remaining calculations unchanged.

The Variational Problem

Given an objective function $J(u)$, we define the optimization problem as

$$\begin{aligned} &\min\ J(u) : \{u \in \mathcal{U}\} \\ &\text{s.t. } \mathbf{X}_u(t) = \mathbf{X}_u(0) + \sum_{j=1}^{\nu} \xi_j \left(\int_0^t h_j\left(\mathbf{X}_u(s)\right) ds \right) \nu_j, \end{aligned} \tag{25}$$

with the ξ_j as independent unit Poisson processes and $\nu_j \in \mathbb{Z}_+^n$ as the molecule change vector associated with reaction j. The dynamic constraint in (25) is a path-wise representation of the perturbation-dependent CTMC $\mathbf{X}_u \equiv \mathbf{X}(u)$ and is commonly referred to as the *random time change model*. Further, we restrict the perturbations to be positive and to fulfill an L^p constraint, i.e.,

$$\mathcal{U} = \{u \in L^p([0,T],\mathbb{R}) \mid u \geqslant 0 \wedge \|u\|_p = E\}. \tag{26}$$

Without further simplifications the variational problem (25) turns out to be analytically intractable. Thus, we assume the perturbation to be a parameterized function, i.e., $u \equiv u(\theta)$ with $\theta \in \mathbb{R}^q$ as a set of q perturbation parameters. In particular, we assume $u(\theta)$ to be an equally spaced, piecewise constant function, with θ specifying the q perturbation levels, i.e.,

$$u(\theta, t) = \sum_{i=1}^{q} \theta_i \mathbf{1}\{t \in \mathcal{T}_i\}, \tag{27}$$

with $\mathcal{T}_i = \{t \in [0,T] \mid (i-1)\Delta \leqslant t < i\Delta\}$ and $\Delta = T/q$. Here we restrict our analysis to the case of $p = 1$, which - in conjunction with the positivity constraint - yields the following discrete optimization problem

$$\min\ J(\theta) : \{\theta \in \mathcal{G}\} \tag{28}$$

with $\mathcal{G} = \left\{\theta \in \mathbb{R}^q \mid \theta \geqslant 0 \wedge \|\theta\|_1 = E\Delta^{-1}\right\}$ as the feasible set. Note that for compactness, the dynamic constraint was omitted in (28).

Stochastic Approximation

In the following, we propose an efficient gradient-based algorithm for numerically minimizing $J(u)$ based on stochastic approximation [115, 126]. Although direct evaluation of $J(\theta)$ is impossible, it is straight forward to obtain noisy estimates of $\widehat{J}(\theta)$ using Monte Carlo integration, such that we can estimate the i-th component of the gradient of $J(\theta)$ as a one-sided finite difference

$$\widehat{\nabla}_i(J) = \frac{\widehat{J}(\theta + h_n e_i) - \widehat{J}(\theta)}{h_n}, \tag{29}$$

with $e_i \in \mathbb{R}^q$ as the i-th canonical base vector and $h_n \in \mathbb{R}$ as the discretization step size. The main idea of the constrained stochastic approximation algorithm is to iteratively update the perturbation parameters as

$$\theta^{n+1} = P\left(\theta^n - \alpha_n \widehat{\nabla}(J)\right), \tag{30}$$

where the sequences α_n and h_n need to be chosen such that $\sum_{n=0}^{\infty} \alpha_n = \infty$, $\sum_{n=0}^{\infty} \alpha_n^2/h_n^2 < \infty$, $\lim_{n\to\infty} \alpha_n = 0$ and $\lim_{n\to\infty} h_n = 0$ to ensure convergence [115]. For all simulations in Section *Simulations*, we choose

$$\alpha_n = \frac{a_0}{A + n^\rho} \text{ and } h_n = \frac{h_0}{n^\gamma},$$

whereas the individual parameters were tuned for each of the problems individually. Function P projects θ back to the nearest point in the feasible region $\mathcal{G}$ (by means of the L^2-metric). In general, such a projection might be tedious to compute. Note however that in our particular case $\mathcal{G}$ defines a canonical simplex in $\mathbb{R}^q$, for which the projection can be solved within a finite number of steps. In this work, we use the algorithm proposed in [156].

Fast Gradient Approximation Using Importance Sampling

Note that the one-sided gradient estimate rests upon $q+1$ Monte Carlo integrations over the path space, which might lead to slow convergence for large q or high-dimensional reaction dynamics. However - such as demonstrated in [195] - the number of required SSA runs can be significantly reduced using importance sampling concepts. We write the perturbed rate constants as $\mathbf{c} \equiv \mathbf{c}(\theta)$. Now, assume that a set of sample paths $\{\mathbf{x}_u^{(i)} \mid i = 1, \ldots, P\}$ has been simulated for a particular θ to obtain an estimate $\widehat{J}(\theta)$. Then, instead of additionally sampling paths for a new parameter set θ' (and corresponding perturbation u'), we can efficiently draw them according to a mixture distribution, i.e.,

$$\begin{aligned} \mathbf{X}_{u'} &\sim \frac{1}{\sum_{i=1}^{P} w_i} \sum_{i=1}^{P} w_i \delta_{\mathbf{x}_u^{(i)}}(\mathbf{x}_{u'}) \\ \text{with } \; w_i &= \frac{p\left(\mathbf{x}_u^{(i)} \mid \mathbf{c}(\theta')\right)}{p\left(\mathbf{x}_u^{(i)} \mid \mathbf{c}(\theta)\right)}, \end{aligned} \tag{31}$$

where $\delta_{\mathbf{x}_u^{(i)}}$ is a Dirac measure over the path space and w_i is referred to as the *importance weight* of sample $\mathbf{x}_u^{(i)}$. Practically, one can sample from (31) by drawing an index i from the discrete distribution defined by the normalized important weights. Then, $\mathbf{x}_{u'}$ is given by the sample path associated with index i, i.e., $\mathbf{x}_u^{(i)}$. In principle, the set of sample paths only needs to be simulated once in order to run the optimization. However - as with all importance sampling techniques - finite sample effects become more significant for larger deviations between θ and θ'. In this work, we simulate new sample paths to obtain $\widehat{J}(\theta)$ and then, use equation (31) to compute $\widehat{J}(\theta + h_n e_i)$ for all $i = 1, \ldots, q$.

Simulations

In the following, we perform simulation studies based on two simple reaction networks. For all the simulations, we assume prior knowledge over the rate constants, e.g., obtained from previous experiments. In particular, we assume a prior of the form

$$\pi(\mathbf{c}) = \prod_{j=1}^{\nu} \Gamma(a_j, b_j)$$

with $a_j = 20c_j$, $b_j = 20$ and c_j as the true parameter. Furthermore - for simplicity - we assume initial conditions to be known. In all case studies, a simple step perturbation was used as a starting point for the numerical optimization. Estimates of the objective $\widehat{J}(u)$ were computed using $40-120$ sample paths.

We studied the algorithm performance under the realistic setting of discrete-time and noisy measurement. We assume a measurement equation in the form of (20), corrupted by additive Gaussian measurement noise with zero-mean and standard deviation $\sigma_Y = 4$. Before we applied the algorithm to a more complicated, nonlinear reaction network, we studied the perturbation design for a simple birth-death process (see Fig. 8), for which the results are easier to interpret. We assume that the birth rate can be controlled by an external perturbation (i.e., $c_1 \equiv u(\theta, t)$), which is optimized such as to minimize the expected logarithm of the posterior variance of c_2.

Figure 8: A simple birth-death process.

Fig. 9 illustrates an exemplary minimization of $\widehat{J}(\theta)$ over the number of update iterations for a three-level input profile applied to the birth-death process. The achieved decrease of the objective function corresponds to roughly two orders of magnitude of the posterior variance.

We performed simulations for the case of one (see Fig. 10A) and two (see Fig. 10B) measurement time points of species A, whose time evolution is denoted $A(t)$. Details on the parameter configuration used in the following simulations are summarized in Table 3.

Parameter	c_1	c_2	q	E
Value	$u(\theta, t)$	0.30	15	200
Unit	s^{-1}	s^{-1}	-	1

Table 3: Parameter configuration for the birth-death model.

Interestingly, high perturbation amplitudes arise immediately before the measurement time points, which we interpret as follows: first - as true

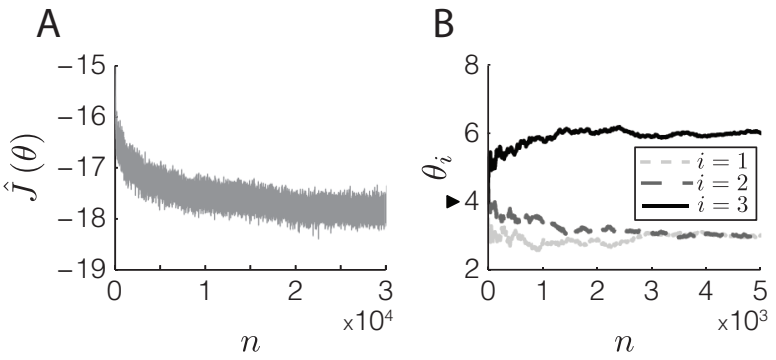

Figure 9: Illustration of the stochastic approximation algorithm. (A) Exemplary minimization of $\widehat{J}(\theta)$ over 30000 update iterations and (B) convergence of a three-dimensional perturbation to the optimum. The perturbation was initialized to a step function, i.e., $\theta_i = 4$ for $i = 1, 2, 3$, indicated by the black triangle.

in the general - perturbations yielding measurements during a dynamic transient are preferable to measurements close to a stationary state. Second - in the particular case of the first order death reaction - strong excitation close before the acquisition time will accumulate many of the events at regions where they - conditional on that excitation - can be inferred or "located" more accurately. It can be seen from Fig. 10 that while the process mean is significantly increased, the standard deviation remains more or less unchanged. In contrast, when considering the step perturbation, most of the transient is missed during acquisition and furthermore, degradation events will be spread over a wider time interval. We also want to point out the significant difference between the perturbations obtained for the incomplete and complete case. For the latter, it was shown in [162] that the expected posterior variance is minimal for the case $\theta_1 = E\Delta^{-1}$ and $\theta_{i\neq 1} = 0$.

We repeated the two-observation experiment for a nonlinear model of transiently induced transcriptional activation. Often cells react to changing environmental conditions, by activating particular transcriptional programs (see e.g., [176]). Sensed at the cell membrane, the stimulus or *stress* is mediated to the nucleus by a translocation of certain transcription factors, which are activated by the signaling cascade. Once in the nucleus, the signaling proteins can initiate transcription of the target genes. After the cell has adapted, the transcription factors relocate to the cytoplasm, giving rise to only a short time period of gene activity. A minimalistic model of the transiently induced transcriptional activation is depicted in Fig. 11, whereas all reactions are modeled according to mass-action kinetics.

We further assume that the intracellular dynamics can be perturbed by means of the rate $c_2 \equiv u(\theta, t)$ (see Table 4 for details). The initial abundances of A, B, AB and C are initialized at 0, 5, 0 and 0 copies, respectively.

We assume that we can obtain noisy measurements of C at two time points with standard deviation $\sigma_Y = 4$. The optimal perturbation was computed for the case of jointly estimating c_1 (degradation) and c_5 (protein

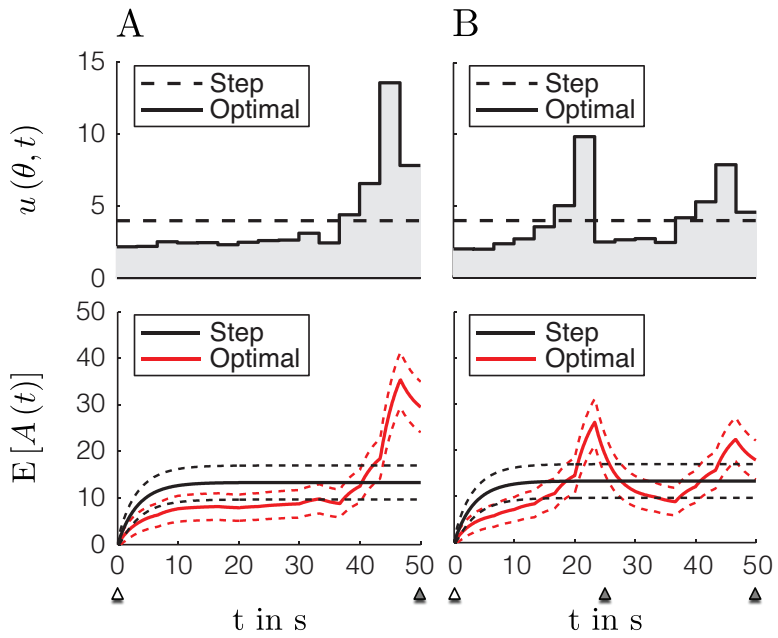

Figure 10: Optimal perturbations and mean process dynamics. Mean dynamics (solid) and the $\pm\sigma$ confidence bounds (dashed) for the step (black) and optimal (red) perturbation were computed by integrating the moment ODEs. The triangles indicate the initial (white) and observation (gray) time points used for perturbation design.

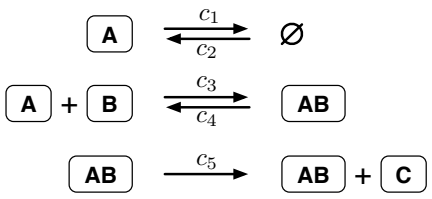

Figure 11: A simple model of transiently induced transcriptional activation. The transient nuclear accumulation of transcription factor (species A) is modeled by production and degradation events. Molecules A can bind to the promoter of the target gene (species B) to form a complex (AB). Transcription of mRNA and translation to the protein (C) is abstracted by a first order production event.

Parameter	c_1	c_2	c_3	c_4	c_5	q	E
Value	0.1	$u(\theta, t)$	0.05	0.20	0.80	15	80
Unit	s^{-1}	s^{-1}	s^{-1}	s^{-1}	s^{-1}	-	1

Table 4: Parameter configuration for the transcriptional model.

synthesis). The resulting perturbation as well as the mean process dynamics of species A and C are depicted in Fig. 12.

Compared to the step response, the optimized perturbation results in a strong initial up-regulation of species A, followed by a period where it

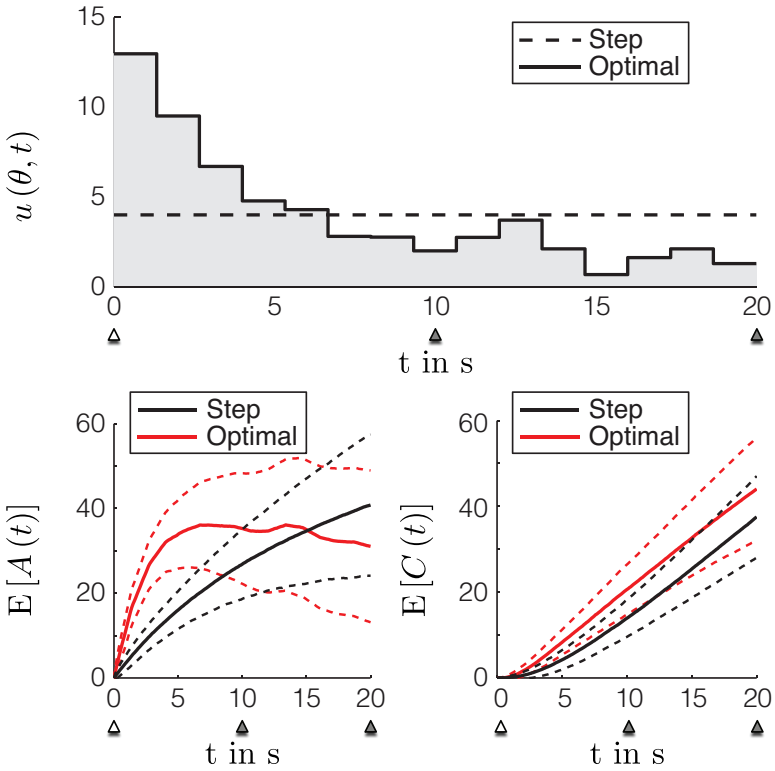

Figure 12: Optimal perturbations and mean process dynamics. Mean dynamics (solid) and the $\pm\sigma$ confidence bounds (dashed) for the step (black) and optimal (red) perturbation were computed over 2000 SSA runs. The triangles indicate the initial (white) and observation (gray) time points used for perturbation design.

decreases again. Intuitively, it seems important to have a high transcription factor abundance during the early time points, such that (a) many degradation events have appeared at the measurement time points, and (b) maximize the temporal window of gene activity, such that many new proteins can be synthesized. This is supported by the increased mean of species C as shown in Fig. 12. However, we want to stress that such explanations cannot be rigorously justified due to the nonlinearity of the reaction network as well as the challenging incomplete data scenario.

Conclusion

We presented a computational framework for the design of optimal experimental perturbations for stochastic reaction kinetics. Based on a CTMC description, we first discussed statistical inference of parameters within the Bayesian framework. Subsequently, we formulated a constrained variational problem in order to render novel experiments *most informative* by means of the expected generalized posterior variance. The analytical complexity of the optimization problem required numerical strategies to solve the variational problem. Here we applied a stochastic approximation algorithm, which was further accelerated using importance sampling concepts.

Simulation studies were performed for the incomplete data scenario using a simple birth-death process and a more complicated nonlinear model of stress-induced transcriptional activation. In all cases, the generalized posterior variance was reduced by orders of magnitude.

Future Work

Although computationally efficient, the proposed optimization algorithm is based on Monte Carlo integrations over a possibly high-dimensional parameter-, path- or measurement space. Especially for large q, this might yield poor convergence, such that either the number of update iterations or the number of Monte Carlo runs must be increased. This issue might be alleviated, by replacing the path-based description by a population-based, e.g., moment-based description of the process distribution [97, 256, 190]. In this case - under certain assumptions - it might be possible to analytically compute gradients of the objective function.

6

MANUSCRIPT 4

REFERENCE

C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, *Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings*, Nature Methods, vol. 11, no. 2, pp. 197-202, doi. 10.1038/nmeth.2794, Jan. 2014.

AUTHOR CONTRIBUTIONS

CZ, MU, MP and HK designed the research. CZ and HK conceived of mathematical methods, performed simulations and analyzed data. MU and SP developed strains. MU and SP performed experiments and measured data. CZ and HK wrote the paper.

SCALABLE INFERENCE OF HETEROGENEOUS REACTION KINETICS FROM POOLED SINGLE-CELL RECORDINGS

Christoph Zechner[1], Michael Unger[1,2], Serge Pelet[3], Matthias Peter[2], and Heinz Koeppl[1,4,5]

ABSTRACT

Mathematical methods combined with measurements of single-cell dynamics provide a means to reconstruct intracellular processes that are only partly or indirectly accessible experimentally. To obtain reliable reconstructions, the pooling of measurements from several cells of a clonal population is mandatory. However, cell-to-cell variability originating from diverse sources poses computational challenges for such process reconstruction. We introduce a scalable Bayesian inference framework that properly accounts for population heterogeneity. The method allows inference of inaccessible molecular states and kinetic parameters; computation of Bayes factors for model selection; and dissection of intrinsic, extrinsic and technical noise. We show how additional single-cell readouts such as morphological features can be included in the analysis. We use the method to reconstruct the expression dynamics of a gene under an inducible promoter in yeast from time-lapse microscopy data.

INTRODUCTION

Statistical inference of unobserved molecular states and parameters that characterize an intracellular process is instrumental for the advance of quantitative biology. Single-cell assays provide particularly informative data to perform this inference. They provide access to the stochastic nature of cellular processes and to the considerable cellular heterogeneity present in even a clonal population of cells.

From the viewpoint of inference, two classes of single-cell data may be distinguished. The first is population snapshot data, provided, for instance, by cytometry techniques [258, 94, 172] or FISH [180, 165]: with such data – when measured over time – any temporal behavior on the single-cell level is necessarily lost. The second class is time-lapse live-cell data [153, 92, 221], wherein individual cells can be followed over time and therefore contain information about the temporal behavior of molecular processes inside a single cell. Evidently, such information is extremely

1 Automatic Control Lab, ETH Zurich, Zurich, Switzerland.
2 Institute of Biochemistry, ETH Zurich, Zurich, Switzerland.
3 Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland.
4 IBM Zurich Research Laboratory, Rueschlikon, Switzerland.
5 Present address: TU Darmstadt, Darmstadt, Germany. Correspondence should be addressed to H.K. (heinz.koeppl@bcs.tu-darmstadt.de)

helpful for inference and renders live-cell measurements superior in this respect. Recently we proposed a complete inference framework for population snapshot data [258]. In particular, cell-to-cell variability was mathematically accounted for, and time-lapse flow cytometry data were exemplarily used to infer kinetic parameters of a gene expression system in yeast. Here we lay out a corresponding inference scheme for time-lapse live-cell data. A very different approach needs to be followed in order to fully use the information contained in such data.

Several inference techniques for stochastic chemical kinetics based on single-cell time-lapse data have been proposed recently [8, 84, 171, 214]. Their focus is largely on inference from observation of a single cell-trajectory because the extension to multiple trajectories is straightforward if one assumes no heterogeneity apart from that caused by the intrinsic noise of chemical reactions [61]. In practice, however, the pooling of several single-cell recordings – which is necessary to obtain reasonable estimates – generates difficulties due to extrinsic contributions to the observed heterogeneity of the cell population [61, 45, 209, 29, 99]: for example, due to differences in cell-cycle stage or cellular translation efficiency. Early attempts to devise inference techniques that account for such heterogeneity had limited scalability with respect to the number of pooled cells [118]. Here we develop a scalable inference scheme, for which the number of unknown parameters is independent of the population size. We used this approach in conjunction with time-lapse microscopy measurements to reconstruct dynamic states and parameters of induced gene expression in yeast.

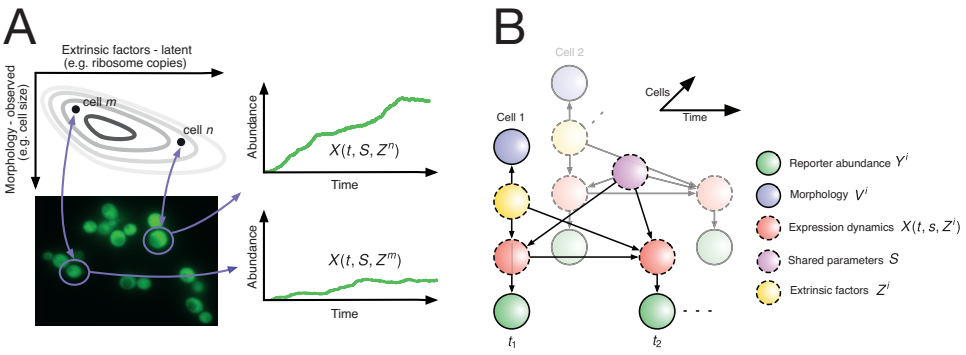

Figure 13: Modeling heterogeneous microscopy data. (A) Schematic generative model of the experimental data. In addition to intrinsic fluctuations, extrinsic factors and their morphological covariates render individual cells different. The gene expression dynamics $X(t, S, Z^i)$ are characterized by a parameter set S that is shared across cells and a set of individual (i.e., extrinsic) parameters Z^i. (B) Corresponding Bayesian mixed-effect model. Nodes denote random variables. Statistical dependency is indicated by directed edges. Nodes with solid borders correspond to experimentally accessible quantities; dashed nodes refer to unobserved, latent variables. Extrinsic factors Z_i are assumed to be drawn from a common distribution.

RESULTS

Inference from heterogeneous live-cell data

We developed a comprehensive and scalable computational framework, Dynamic Prior Propagation (DPP), for the inference of stochastic biochemical processes from pooled single-cell time-lapse measurements. The approach is based on a hierarchical Markov model (Fig. 13), accounting for cellular heterogeneity caused by intrinsic molecular fluctuations and extrinsic [61, 45] contributions. We model the latter through parameters that are believed to vary across pooled cells: for example, the protein translation rate. We refer to those parameters as 'extrinsic factors' and assume them to be time invariant throughout. Dynamically changing extrinsic factors [29, 99] complicate the inference problem considerably, and no practical solution is known to date. We assume that only a very limited number of the molecular species (i.e., only one in the case studies below) can be measured at discrete time points and that the measurements are corrupted by noise. For every inferred quantity, the framework returns a probability distribution characterizing how well this quantity is determined by the acquired data (posterior distribution).

In a straightforward attempt at inference, the number of unknown parameters would increase with the number of pooled cells [118], making inference computationally very demanding for even modest population sizes. In contrast, DPP relies on a particular marginalization [1] of the biochemical process with respect to extrinsic factors, yielding a single dynamic model that represents the heterogeneous cell population in an exact manner. Instead of depending on the extrinsic factors of a cell, the resulting marginal process directly depends on how those extrinsic factors are distributed across the population. For instance, in the case of a randomly distributed translation rate, the process depends directly on the parameters of that distribution, such as the mean and variance (we refer to those parameters as 'extrinsic statistics'). The presented framework combines the marginal process with sequential Monte Carlo techniques [54] to yield a scalable inference algorithm. Image-based single-cell techniques can additionally capture morphological features of cells such as their volume or shape (Fig. 13). Such features have been shown to correlate well with extrinsic factors [209, 187]; in contrast to previous approaches (such as size normalization), DPP offers a principled way to leverage such additional information for inference. As with the extrinsic factors, we characterize morphological features by a distribution and subsequently infer the parameters of this distribution from data (*Online Methods* and *Supplementary Note 1*).

Application to simulated gene expression data

We first studied the proposed inference framework using simulated data of a simple two-state gene expression model [180] under realistic measurement conditions (Fig. 14 A). We assume that the target gene or promoter can be activated through an exogenous signal: for example, the binding of a transcription factor.

We simulated extrinsic variability by introducing a gamma-distributed variability in the protein translation rate. We collected simulated data on 20 cells, on which we applied DPP using 10,000 Monte Carlo samples per measurement time instance (Fig. 14 A,B). We inferred posterior distributions over kinetic parameters, states, extrinsic statistics and an acquisition noise parameter that characterizes the measurement uncertainty. (Fig. 14 B and *Supplementary Note 2*). Furthermore, we inferred states with respect to mRNA and protein levels (Fig. 14 C) in two exemplary cells with different translation efficiencies.

The inferred posterior distribution over unobserved states can also be used to reconstruct promoter activation and transcription events (Fig. 15). We simulated a double-pulsed induction of gene expression with the model described in Figure 14. We used noisy versions of the simulated protein abundance at sparse time points as our available measurements and reconstructed mRNA and promoter dynamics using DPP. In general, the inverse problem of reconstructing promoter activation states from slow protein dynamics is considerably ill posed. However, we find that accurate detection of promoter states is indeed possible within the considered scenario (Fig. 15). We note that reconstructing a promoter activation sequence by simply determining the maximum of the posterior distribution over promoter states (Fig. 15) at each time point does not yield the desired activation sequence.

Application to experimental gene expression data

We used DPP to reconstruct the expression dynamics of an artificially controlled gene expression system in Saccharomyces cerevisiae based on hormone-dependent activation of the chimeric transcription factor GAL4DBD.ER.VP16 (GEV) [142, 149]. GEV consists of a strong transcriptional activator, made by fusing the GAL4 DNA-binding domain (GAL4DBD) with the hormone-binding domain of the human Estrogen Receptor (ER) [142] and the transcription-activating domain of the herpes simplex virus protein, VP16 [192]. In its inactive state, GEV associates with the Hsp90 chaperone complex and resides in the cytoplasm. Upon addition of β-estradiol to the extracellular medium, Hsp90 disassociates from the complex, and active GEV translocates to the nucleus, where it activates transcription of genes under a GAL1 promoter.

We engineered a strain that allows a combined readout of GEV translocation and β-estradiol-induced gene expression. A GEV-mCherry construct

in combination with a nuclear marker allows computation of the ratio of nuclear to cytoplasmic GEV. The strain also carries a destabilized [236, 89] version of the Venus fluorescent protein (Y-Venus) under control of a GAL1 promoter (*Online Methods* and *Supplementary Note 3*).

We carried out fluorescence microscopy using a flow chamber for rapid media exchange. We exposed the cells to a 30 min pulse of 50 nM β-estradiol and analyzed [175] the movies generated from the time-lapse images to quantify GEV-mCherry nuclear localization and Y-Venus reporter gene expression in individual cells.

We performed calibration experiments with reference strains to map the recorded fluorescence intensities to total protein abundances (*Supplementary Note 4*). On the basis of a one-sided Kolmogorov-Smirnov test, we corrected for the time delay in protein measurements that arises from unmodeled sequential events such as mRNA export, post-transcriptional or post-translational modifications and reporter maturation (*Supplementary Note 4*). We used 20 single-cell trajectories of Y-Venus abundance for subsequent analyses. We used the average translocation curve as input to our gene expression models, basing these on the assumption that translocation occurs uniformly across cells and given the high abundance of GEV.

Modeling pGAL1 Y-Venus expression

We investigated three different models of eukaryotic gene expression and determined how well they are supported by our experimental data using Bayesian model selection. In addition to the canonical two-state model [180] (Fig. 13 A and Supplementary Fig. 19 A) of a promoter, we considered a three-state model [26] wherein initiation-complex assembly is followed by a slow activation step representing either RNA polymerase binding or chromatin remodeling (Supplementary Fig. 19 B) and a three-state model including a refractory state [92, 221] (Supplementary Fig. 19 C). We repeated the model selection analysis multiple times to check the robustness of the obtained ranking (*Supplementary Note 5*). The two-state model consistently ranked best, and it was closely followed by the three-state model with a refractory state. We also performed a model selection for two competing models of measurement noise (i.e., normal and log normal) and found strong evidence for log-normally distributed noise (*Supplementary Note 5*).

Modeling predicts mild bursting in the GEV-pGAL1 system

We used the experimental data to perform parameter inference, state reconstruction and promoter activity detection as described above for the synthetic case study (Fig. 16 B). Our model estimates an mRNA half-life of around 10 min and mRNA synthesis rate of six molecules per minute, results in line with previous findings [260]; the latter value is above most reported rates for constitutively expressed genes [260], which appears con-

sistent with our use of the strong VP16 activator. This synthesis rate, together with the length of 850 bp for the Y-Venus protein and a reported elongation speed of 2 kb/min [148] for GAL-driven genes, indicates that there need to be roughly three RNA polymerases on average on the gene. We reconstructed states for two cells with different Y-Venus abundance (Fig. 17 A). The predicted timing statistics of the promoter activation sequences indicate that, for successful initiations, around 2.5 transcripts per active promoter state are produced on average, suggesting that transcription reinitiation, and thus mild bursting, takes place in this expression system.

We performed additional experiments with 25 nM and 100 nM β-estradiol (Supplementary Figs. 20 and 15) and validated the inferences of the calibrated model against these experimental results. The model predictions agree well with the experimentally obtained data across different concentrations of β-estradiol (Fig. 17 B). Further discussion of the obtained results and their comparison with other analytical and experimental work can be found in *Supplementary Note 6*.

Noise contribution in pGAL Y-Venus expression

Our model can also indicate to what extent a cell's expression level is explained by extrinsic and intrinsic factors (Fig. 17 A). Although two cells might show similar mRNA levels, one may express substantially more Y-Venus owing to a higher translation rate.

To test this further, we forward-simulated the inferred model to quantify the different sources of variability in the measured reporter abundance using the law of total variance (Online Methods). More specifically, we separated intrinsic, extrinsic and technical contributions to the overall variability (Fig. 18 A). We note that any systematic bias to the technical contribution – for instance, by the image segmentation algorithm – is not considered. The inferred model predicts that the variability in pGAL1 Y-Venus expression is substantially driven by extrinsic factors. This is further supported by the fact that a model that accounts only for intrinsic and technical noise fails to predict the cell-to-cell variability in the data (*Supplementary Note 7*).

We validated our predictions experimentally using an independent data set consisting of a dual-reporter (YFP-CFP) readout of the same promoter under identical experimental conditions (Supplementary Fig. 22). We quantified intrinsic and extrinsic noise using a conventional approach [182]. We note that this approach does not account for technical variability (such as that due to image segmentation errors), which will hence be subsumed in the intrinsic and/or extrinsic parts. Our predicted noise contributions are in good agreement with the variance decomposition from the dual-reporter experiment across different concentrations of β-estradiol (Fig. 18 B) and across time points (Fig. 18 C and Supplementary Fig. 23). Although the overall noise characteristics are well captured by the model predictions,

deviations are visible at early time points. Although background fluorescence levels have been estimated and subtracted from the dual-reporter data, correlated residuals will persist owing to estimation uncertainties. Hence, in the case of very low abundances (i.e., at early time points), such residuals are likely to dominate and cause an overestimation of extrinsic contributions.

Morphological features and extrinsic variability

In addition to making use of protein measurements, DPP allows both incorporation of additional single-cell readouts such as morphological features and quantification of statistical dependencies between such readouts and a population's extrinsic factors.

We hypothesized a dependency between volume increase during the observation time interval and the translation efficiency and quantified it using DPP (Fig. 18 D). Consistent with previous studies [45], we found that volume increase positively correlates with translation efficiency but that it does not explain all extrinsic variability in the intensity trajectories. This provides evidence for the fact that simple normalization through morphological features (for example, forward scattering in flow cytometry data [258]) cannot sufficiently correct for extrinsic variability in gene expression data.

DISCUSSION

Inference of stochastic dynamics of a cellular process from live-cell recordings of only one cell is inherently ill posed. Pooling even a few single-cell recordings substantially improves inference accuracy (*Supplementary Note 8*). However, the large degree of extrinsic variability in such data adds complexity, and straightforward inference approaches cannot perform well with the added dimensionality. Our method rests on a recursive inference scheme, whose strength is achieved by marginalizing the dynamics of the process being studied over extrinsic factors and uncertain kinetic parameters. This has the following consequences. First, kinetic parameters are no longer sampled along with the dynamic states, which results in an improved accuracy of the desired posterior statistics (*Supplementary Note 9*). Second, in the context of cell-to-cell variability, the marginalization yields better scalability with respect to the number of pooled cells and is therefore advantageous over previous approaches that require intermediate sampling of extrinsic factors [118].

Our approach is related to the Rao-Blackwellized particle filter [54], in which a set of linear dynamic states is marginalized out to reduce the dimensionality. It applies to any nonlinear reaction network with linearly parameterized propensities, such as those obtained by mass-action principles. We demonstrated the validity of the method in a simulation study of a two-state gene expression model (Figs. 14 and 15).

When applied to an engineered β-estradiol-induced gene expression system in yeast, DPP together with the performed protein calibration measurements offers absolute estimates of kinetic parameters, unobserved molecular states and population heterogeneity. Hence, it allows the inference of several quantities and their uncertainty without dedicated experimental techniques for each quantity [180, 260].

Our results are based on the two-state gene expression model because Bayesian model selection indicated less supportive evidence for more complex models. In particular, no refractory promoter state was evidenced by the data.

By pooling heterogeneous single-cell recordings, we used our model to dissect the different contributions to cell-to-cell variability, which traditionally requires experiments such as two-color assays [61, 182]. In line with previous studies on GAL-driven genes [26, 182] and with orthogonal dual-reporter data, we found that extrinsic noise is a substantial source of cell-to-cell variability for such genes. Moreover, we observed a characteristic decrease of the intrinsic noise components accompanying the increase in mean expression level over the course of induction, a pattern consistent with Poissonian noise.

We see clear evidence in favor of log-normally distributed noise, a result indicating that measurement errors scale with the measured fluorescence intensity. Incorporating morphological features [187] in addition to fluorescence data can increase the predictive power of computational models with respect to extrinsic factors. The statistical dependency between translation efficiency and volume increase extracted by the algorithm is coherent with earlier findings [45] that both quantities are positively correlated.

ACKNOWLEDGMENTS We want to thank H.R. Kuensch and J. Hasenauer for their valuable feedback on the manuscript and O. Aalen for providing us with his technical report from 1988. We thank F. Rudolf for help in designing and cloning the Y-Venus destabilized reporter and S. Lee with the fluidic setup. C.Z., M.U. and H.K. acknowledge support from the Swiss National Science Foundation, grant no. PP00P2_128503 and SystemsX.ch. S.P. and M.P. acknowledge support from the European project UNICELLSYS, European Research Council, SystemsX.ch organization (LiverX), Swiss National Science Foundation and ETH Zurich. M.U. receives support from the Life Science Zurich PhD Program on Systems Biology of Complex Diseases; and M.U., M.P. and H.K. acknowledge support from the Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland.

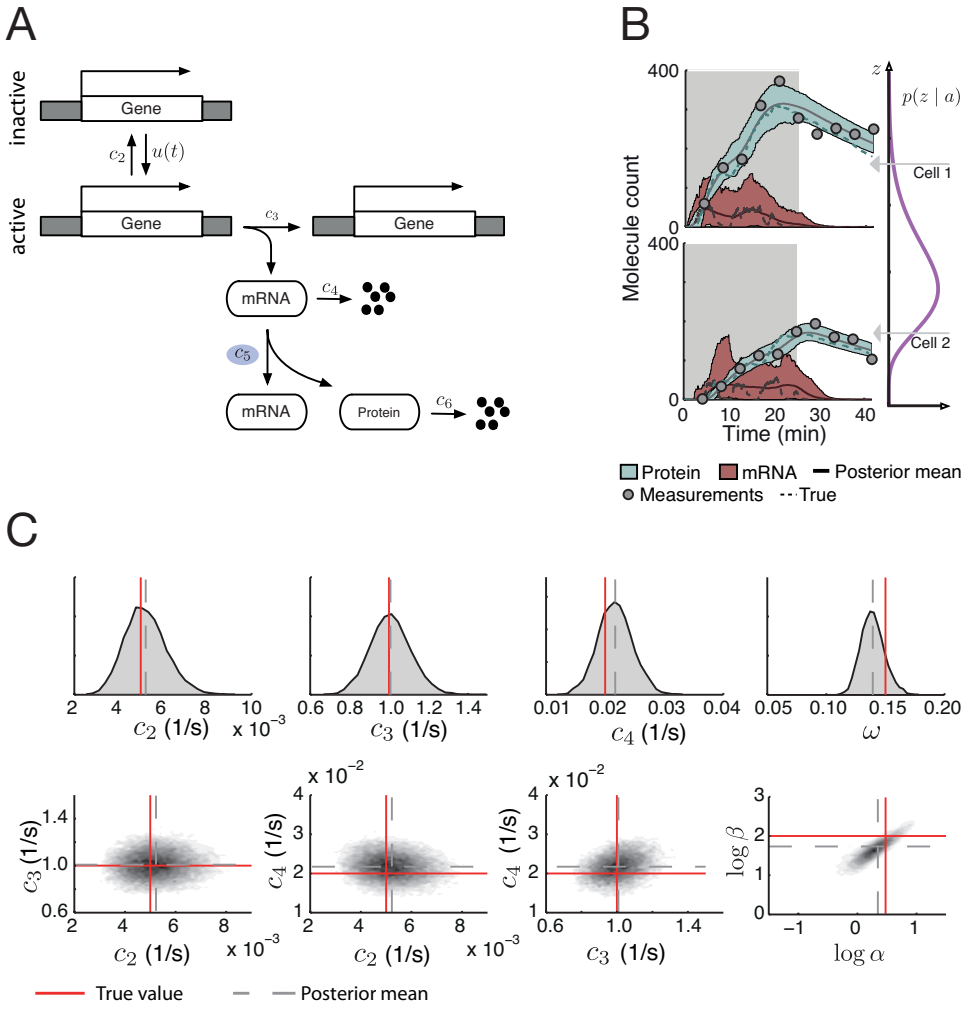

Figure 14: Parameter and state inference using simulated measurements. (A) Schematic two-state gene expression model. All reactions are modeled according to mass action. The model comprises four species and six reactions associated with kinetic parameters $c_1, \ldots, c_6$. The gene activation event is controlled by a time-varying rate, i.e., $c_1 = u(t)$. We assume a gamma-distributed heterogeneity in the translation efficiency, i.e., c_5 drawn from a gamma distribution $G(\alpha, \beta)$ with the extrinsic statistics α and β. (B) Parameter inference from simulated protein measurements using 20 cells. The plots show inference results for the three kinetic parameters (c_2, c_3, c_4), the extrinsic statistics α and β and the scaling parameter ω of the log-normal acquisition noise. Two-dimensional posterior density plots are used to visualize the a posteriori correlations between pairs of parameters. DPP was performed using 10,000 samples per time instance. (C) Inferred mRNA and protein abundance. The 5 and 95 percentiles of the inferred state distributions and their mean values are shown for two representative cells with different translation efficiencies; the gray shaded area indicates the window of induction. The thick violet line illustrates the assumed gamma distribution $p(z \mid a)$ over the extrinsic factor, i.e., the translation rate.

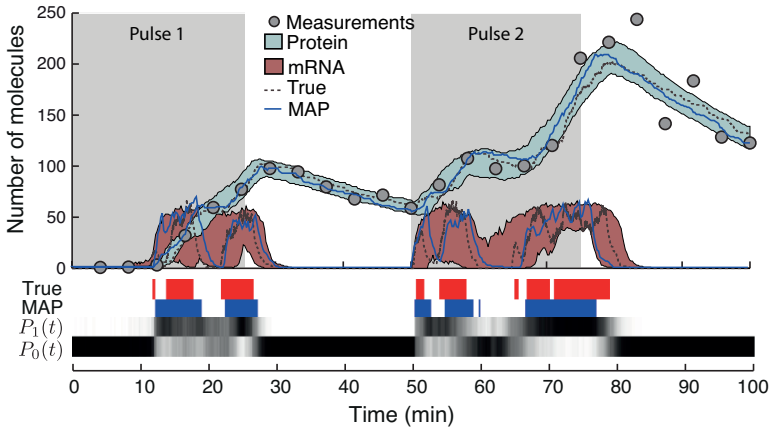

Figure 15: Reconstructed gene expression and promoter dynamics using simulated measurements. Measurements (circles) were obtained by simulating two induction pulses (gray boxes). The lines bounding shaded areas denote 5 and 95 percentiles of the posterior distributions over protein (cyan) and mRNA (red) dynamics; true trajectories (dashed) and inferred MAP trajectory (blue) are indicated by lines. Also shown are the posterior probabilities for the promoter to be active or inactive, $P_1(t)$ and $P_0(t)$, respectively; black denotes a probability of 1. The two upper rows indicate the true (red) and inferred MAP (blue) promoter activation sequences.

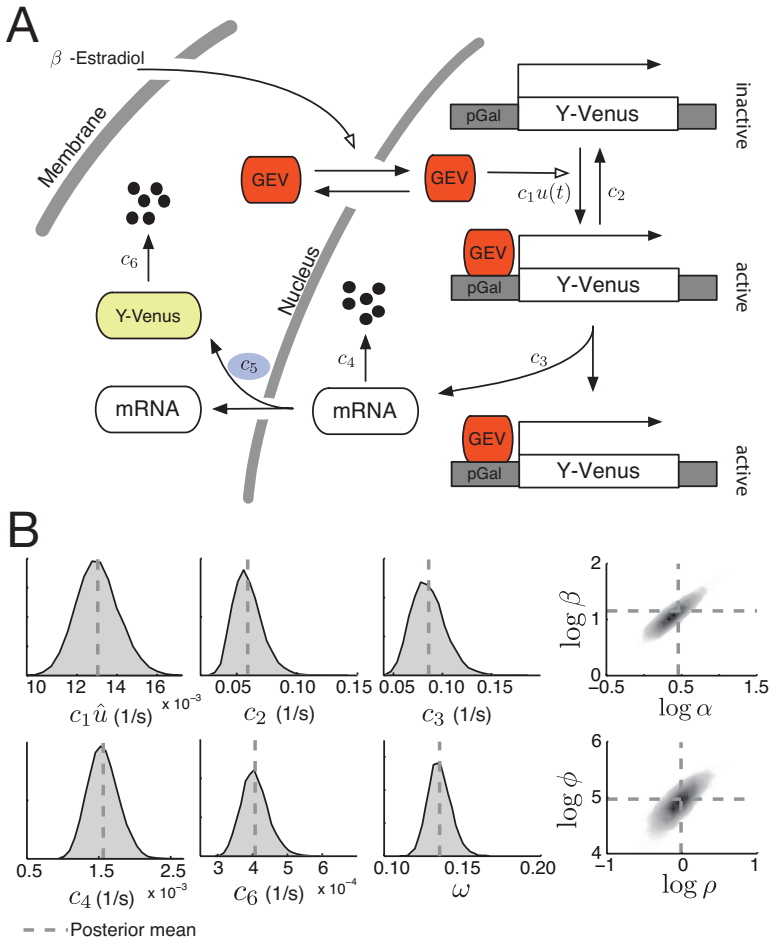

Figure 16: Modeling and parameter inference for induced gene expression in yeast. (A) Stochastic mass-action model with extrinsic noise introduced by a variable translation efficiency c_5 (blue). (B) DPP was performed using 10,000 and 20,000 samples for the first and subsequent iterations, respectively. The plots show the posterior distributions over all kinetic parameters $(c_1, \ldots, c_4, c_6)$, the acquisition noise parameter ω, the extrinsic statistics (α, β) and the morphological shape parameters (ρ, ψ). The marginal posterior for the gene on-rate is shown for $c_1\hat{u}$, where $\hat{u}$ is the temporal average over the modulating GEV intensity.

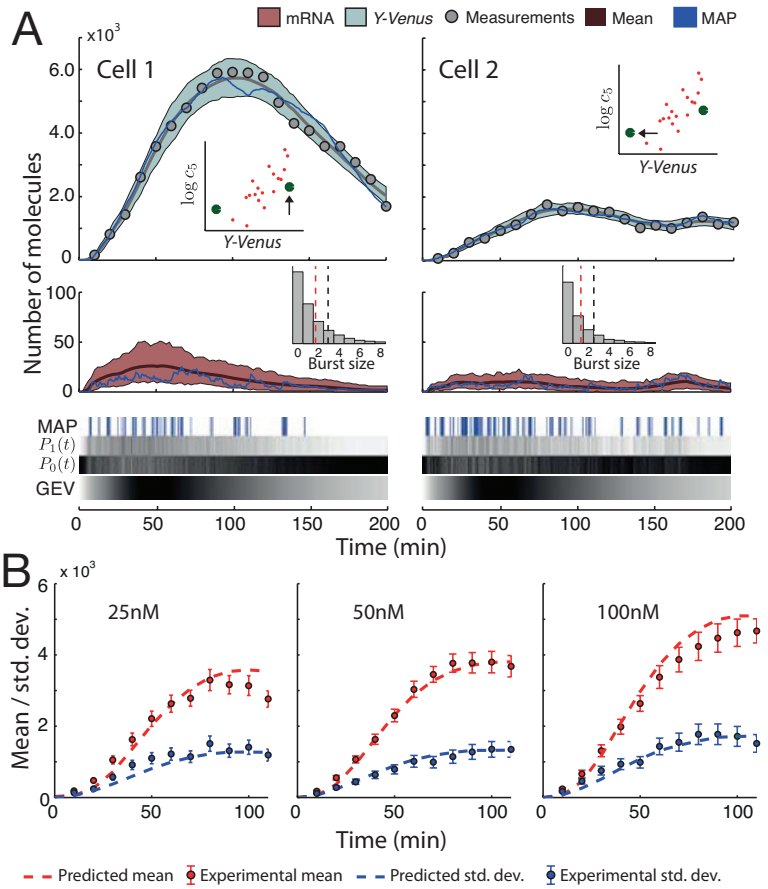

Figure 17: State reconstruction of heterogeneous reporter dynamics. (A) Inferred dynamics for two cells with different Y-Venus abundance. GEV induction over time is shown as an intensity map, where white and black coloring denote minimal and maximal abundance, respectively. MAP, maximum a posteriori. The inset scatter plots indicate the inferred expression regime, where each dot represents a single cell and the arrows point toward the respective cell; the inset x and y axes correspond to the logarithm of the temporal mean of Y-Venus and the mean posterior estimate of translation rate, c_5, respectively. The subpanels below show the inferred mRNA and promoter dynamics. Therein, the insets show the distribution of transcripts per on cycle (i.e., burst size) over all posterior promoter activation sequences, with a mean of 1.5 (red dashed lines) and an adjusted mean of 2.5 when we took only successful initiation events into account (black dashed lines). The lines bounding shaded areas denote 5 and 95 percentiles. (B) Validation of predicted reporter dynamics. Y-Venus trajectories were recorded upon application of β-estradiol pulses of 25, 50 and 100 nM. The plots show a comparison between the predicted and experimental means and s.d. of the Y-Venus dynamics. The error bars indicate the standard errors of the experimentally obtained quantities.

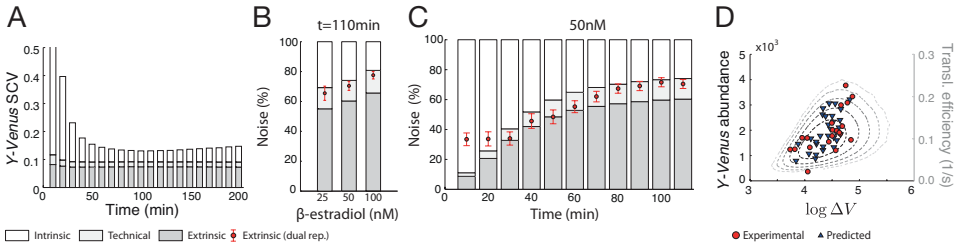

Figure 18: Sources of cell-to-cell variability in reporter expression. (A) The inferred model was used to compute the SCV of the Y-Venus abundance. The total SCV was decomposed into technical, intrinsic and extrinsic components (Online Methods). (B, C) A dual-reporter data set was recorded (red) for the same induction system and experimental conditions as in Figure 17 B. Intrinsic and extrinsic noise contributions were estimated [182]. Error bars indicate standard errors of the experimentally obtained quantities ($n \approx 150 - 300$). (B) Comparison between dual-reporter experiments and model predictions at a fixed time point (110 min) across different concentrations of β-estradiol. (C) Noise decomposition across different time points for the 50 nM β-estradiol pulse experiment. (D) Dependency between volume (V) increase and Y-Venus abundance. Intensity trajectories, volume increase and translation efficiency (c_5) were computed via forward simulation of the inferred model. The plot shows the Y-Venus abundance at 200 min versus the volume increase from 0 min until 200 min for the predicted (triangles) and experimental data (circles). The inferred statistical dependency between the volume increase and the translation (transl.) efficiency is indicated by the dashed iso-lines.

ONLINE METHODS

Fluorescence Microscopy

All experiments were performed on the same epifluorescence microscope (Eclipse Ti, Nikon Instruments), 60x (NA 1.4) oil objective and specific (CFP/YFP/mCherry) excitation and emission filters located in an incubation chamber set to maintain 30 °C. Imaging conditions and parameters were kept constant for all experiments. For each time point, images were acquired at multiple positions using a motorized xy stage, and the focal plane was maintained using a Nikon Perfect Focus System. The microscope and peripheral hardware was computer controlled using μManager [58]. Respective cell chambers were treated with filtered solution of concanavalin A dissolved in PBS (1 mg ml^{-1}) for 30 min and were subsequently rinsed with PBS. Single colonies of the respective yeast strain were picked, inoculated in synthetic (SD) medium and grown overnight at 30 °C. The saturated cultures were then diluted and grown in log phase for at least two doubling times (>4 h). Before they were loaded into the imaging chambers, the cell suspensions were diluted again (OD600 = 0.01) and briefly sonicated.

Pulse Experiments

Single-cell traces were recorded by fluorescence microscopy with a 30 min induction pulse of 25, 50 and 100 nM β-estradiol. The pulses were done by switching between two hydrostatic-pressure (1 PSI) driven flows (SD-full and SD-full + β-estradiol) using a three-way solenoid valve (The Lee Company) connected to the cell chamber (μ-Slide $VI^{0.4}$, Ibidi). Owing to the mutually exclusive switching of the input sources, a constant flow through the cell chamber could be achieved. The valve was connected to a computer via a USB interface board (National Instruments) and controlled using custom software (Matlab, MathWorks). All connections were done using PTFE tubing.

Image Analysis

All microscopy images were analyzed with the YeastQuant platform [175]. The GEV relocation and Venus expression time-lapse movies were segmented on the basis of the nuclear CFP image from the HTA2-CFP marker. The cell boundary was detected as a secondary object surrounding the nucleus on the basis of the mCherry image. The expression of the Y-Venus protein was quantified as the total intensity in the cell. The nuclear accumulation of GEV-mCherry was quantified as the ratio between the average intensity within the nucleus and the cytoplasm, respectively. The expression levels of the YFP-tagged proteins were measured with illumination conditions similar to those used for the Y-Venus imaging. The cells were

segmented on the basis of two bright-field images, and the total cellular intensity of the cell was calculated.

Dual-Reporter Experiments

The intrinsic and extrinsic noise of the GEV system was characterized in cells bearing pGAL1-quadrupleVenus and pGAL1-quadrupleCFP reporters. The GEV was expressed under the control of the constitutive ADH promoter. We used the same experimental protocol as for the single-reporter experiments. The bright-field images were used to segment the cells, and the total intensity of the cell in the YFP and CFP channel was quantified for 150-300 cells per condition (Supplementary Fig. 4). The intrinsic and extrinsic noise was calculated as described in ref. [182]. The time-resolved noise decompositions for the 25 nM and 100 nM β-estradiol pulse experiments are shown in Supplementary Figure 5.

Stochastic Modeling

We consider a well-mixed reaction system with d chemical species and ν reaction channels. The random state vector $X(t)$ collects the copy numbers of species at time t representing, for instance, copy numbers of mRNA and protein. We denote by $\mathbf{X}$ a whole trajectory of $X(t)$ over the time interval $[0, t]$. Subsequently, we will use the convention of denoting a random quantity by an upper case letter and its realization by the corresponding lower case letter; this also applies to Greek letters. The propensity functions of the reactions are assumed to be of the form $h_j(x) = c_j g_j(x)$, with c_j representing the stochastic rate constant and g_j a function of the current state x. We denote the set of all rate parameters as $C = (C1, \ldots, C_\nu)$ and assume that they can be split into a set Z of extrinsic factors (such as translation rate) that can vary across cells and a set S that is shared among clonal cells (for example, elementary dissociation rate). Assuming time invariance of extrinsic factors Z, we assign to them a probability distribution $p(z \mid a)$, where the extrinsic statistics a specify the shape of this distribution and hence the population's heterogeneity. Given a population of M cells, the mth cell's state is then described by a conditional continuous-time Markov chain $X^m \mid (Z^m, S)$.

Modeling the Measured Data

Experimentally we can retrieve noisy measurements of a few molecular species for M cells at different measurement times t_l with $l = 1, \ldots, N$. The acquisition error associated with the experimental technique is characterized by a conditional measurement density $p(y^m \mid x_l^m, \omega)$ with $x_l^m = x^m(t_l)$ and ω as the realization of an unknown distribution parameter Ω such as the acquisition-noise variance. Furthermore, we define the state trajectory of cell m between the lth and the kth measurement time as $\mathbf{X}_{l:k^m}$

and denote by $Y_{l:k^m}$ the corresponding set of measurements. Note that in contrast to $Y_{l:k^m}$, the state trajectory $\mathbf{X}_{l:k^m}$ denotes a continuous-time sample path between t_l and t_k. Morphological features (such as cell volume) are incorporated by introducing morphological covariates V^m and hypothesizing a statistical dependency between these covariates and the extrinsic factors Z^m. This dependency is described by a conditional density $p(v^m \mid z^m, b)$, with b a set of shape parameters characterizing this conditional density.

Heterogeneous Kinetics

The presence of extrinsic factors causes an increase of the parameter dimension with the number M of pooled cells. To overcome this problem, we marginalize the trajectories over extrinsic factors

$$p(\mathbf{X} \mid s, a) = \int p(\mathbf{X} \mid z, s)\, p(z \mid a)\, dz \tag{32}$$

giving rise to a marginal process $X \mid (S, A)$ that directly depends on the extrinsic statistics A. To illustrate its construction, let us assume that the jth component of C is the only extrinsic factor in the model. On the basis of the innovation theorem for counting processes [1], we can show that $X \mid (S, A)$ is again a jump process, where the propensity corresponding to the extrinsic parameter can be generally written as

$$h_j(\mathbf{x}, t) = E[Z \mid \mathbf{x}, a]\, g_j(x(t)) \tag{33}$$

where $E[Z \mid \mathbf{x}, a]$ denotes the conditional expectation of Z given a complete trajectory x and the extrinsic statistics a. Convenient analytical evaluation of equation (33) depends on the distributional assumption of Z.

For instance, the gamma distribution forms a reasonable compromise between analytical tractability and flexibility. More specifically, it represents a very versatile distribution on the positive orthant, ranging from overdispersed and right-tailed to underdispersed and symmetric distributions. Importantly, it was also shown to be well justified in the context of stochastic chemical kinetics [69, 222].

If Z follows a gamma distribution, i.e., $Z \mid a \sim G(\alpha, \beta)$ with $a = (\alpha, \beta)$, we have that [245, 124]

$$Z \mid (\mathbf{x}, a) \sim G\left(\alpha + r_j, \beta + \int_0^t g_j(x(\tau))d\tau\right) \tag{34}$$

and hence

$$h_j(\mathbf{x}, t) = \frac{\alpha + r_j}{\beta + \int_0^t g_j(x(\tau))d\tau} g_j(x(t)) \tag{35}$$

where r_j denotes the number of occurrences of reaction j in x. As a consequence of this marginalization, the Markov property is lost. However, when augmenting the state space by the summary statistics

$$T(\mathbf{x}) = \left(r_j, \int_0^t g_j(x(\tau))\, d\tau\right) \tag{36}$$

of the path $\mathbf{x}$, the Markov property is recovered and, hence, stochastic simulation can be efficiently performed using available methods [9]. Mathematical proofs and derivations for the general multivariate case, the extension to morphological covariates and a simulation algorithm are given in Supplementary Note 1.

Statistical Inference Algorithm

As a general consequence of the above marginalization, it follows that any kind of parametric uncertainty linearly entering the propensities can be integrated out and directly encoded into the process dynamics. This seems generally useful from a Bayesian viewpoint, where parameters are characterized by prior uncertainty. For instance, if the ith rate parameter C_i is associated with a known prior distribution, for example, $C_i \sim G(\alpha_i, \beta_i)$, the marginal propensities are obtained in analogy to equation (33). Marginalization with respect to every kinetic parameter in conjunction with a sequential Monte Carlo techniques leads to the proposed method, termed Dynamic Prior Propagation (DPP). The goal of DPP is to compute the marginal posterior distribution

$$\begin{aligned} &p\left(\mathbf{x}_{1:N}^1, \ldots, \mathbf{x}_{1:N}^M, a, b, \omega \mid \left(y_{1:N}^1, \nu^1\right), \ldots, \left(y_{1:N}^M, \nu^M\right)\right) \propto \\ &\left[\prod_{m=1}^{M}\left(\prod_{l=1}^{N} p\left(y_l^m \mid x_l^m, \omega\right)\right) p\left(\nu^m \mid, b, a\right)\right] \\ &\times p\left(\mathbf{x}_{1:N}^1, \ldots, \mathbf{x}_{1:N}^M \mid a, b, \nu^1, \ldots, \nu^M\right) p(a) p(b) p(\omega) \end{aligned} \tag{37}$$

where $(y_{1:N^m}, \nu^m)$ denotes the tuple of measurements available for the mth cell.

According to a Bayesian filtering approach, the posterior distribution can be determined recursively over time. Consequently, the original sampling problem breaks up into a sequence of subproblems with reduced dimensionality.

Recursive sampling approaches inherently suffer from sample degeneracy as soon as constant parameters are estimated in addition to the dynamic states. Although a majority of the parameters are integrated out, A, B and Ω remain in the model. A standard approach to avoid such degeneracies is to apply an invariant kernel to the static parameters at each time instance, diversifying the parameter samples [219]. Here we use a Metropolis-within-Gibbs Markov Chain Monte Carlo (MCMC) scheme, where the latent space is further divided into blocks, which are resampled successively. This requires sampling from the full conditional distributions, which can be determined using the notion of Markov blankets [119]. A full description of the algorithm and different variants thereof can be found in Supplementary Note 1.

Note that posterior distributions over $Z^1, \ldots, Z^M$ and S are not directly computed by the marginalized inference scheme. However, they can easily be reconstructed via the law of conditional probability, as described

in Supplementary Note 1. Empirical model evidences and Bayes factors are directly calculated by the algorithm and do not require further computations; the corresponding equations and their derivations are given in Supplementary Note 1. A Matlab toolbox for DPP with a detailed tutorial and a simple graphical user interface is made available at https://github.com/koeppllab/DPP/.

Model-Based Noise Decomposition

The law of total variance is applied to dissect the total variability into intrinsic, extrinsic and technical contributions. In particular, it holds that

$$\begin{aligned} \mathrm{SCV}\left[Y_l\right] = & \underbrace{\frac{\mathrm{E}\left[\mathrm{E}\left[\mathrm{Var}\left[Y_l \mid X_l\right] \mid Z\right]\right]}{\mathrm{E}\left[Y_l\right]^2}}_{\text{technical}} \\ & + \underbrace{\frac{\mathrm{E}\left[\mathrm{E}\left[\mathrm{Var}\left[Y_l \mid X_l\right] \mid Z\right]\right]}{\mathrm{E}\left[Y_l\right]^2}}_{\text{intrinsic}} + \underbrace{\frac{\mathrm{E}\left[\mathrm{E}\left[\mathrm{Var}\left[Y_l \mid X_l\right] \mid Z\right]\right]}{\mathrm{E}\left[Y_l\right]^2}}_{\text{extrinsic}} \end{aligned} \tag{38}$$

For the decomposition, the model parameters were set to their mean posterior values, and the individual quantities were obtained via forward simulation. For the comparison with the dual-reporter experiments (Fig. 6 and Supplementary Fig. 5), the degradation rate of the model was set to the previously reported half-life of the quadruple-Venus reporter [258] in order to account for differences in the reporter lifetimes. A derivation of equation (38) can be found in *Supplementary Note 1*.

SUPPLEMENTARY FIGURES

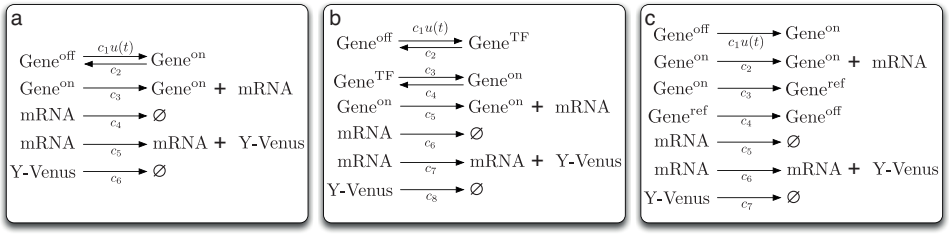

Figure 19: Three candidate models for β-estradiol-induced gene expression. (a) Two-state gene expression model. (b) Three-state gene expression model where initiation-complex assembly is followed by a slow activation step representing either RNA polymerase (RNAP) binding or chromatin remodeling. (c) Three-state model with refractory state.

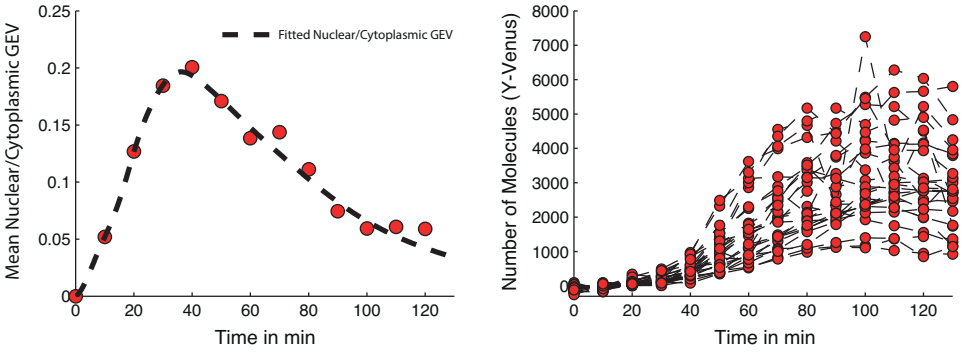

Figure 20: Calibrated single cell traces of Y-Venus expression over time for a 25 nM pulse of β-estradiol.

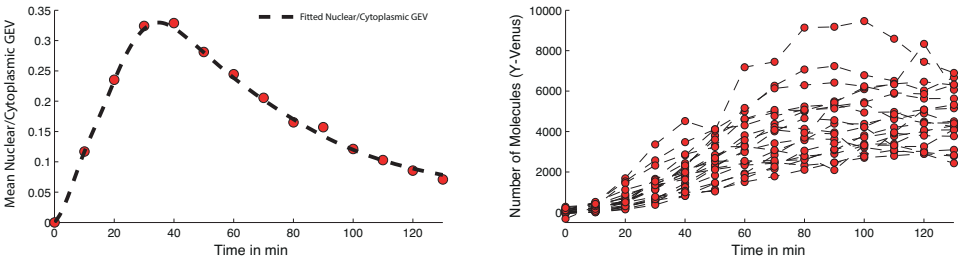

Figure 21: Calibrated single cell traces of Y-Venus expression over time for a 100 nM pulse of β-estradiol.

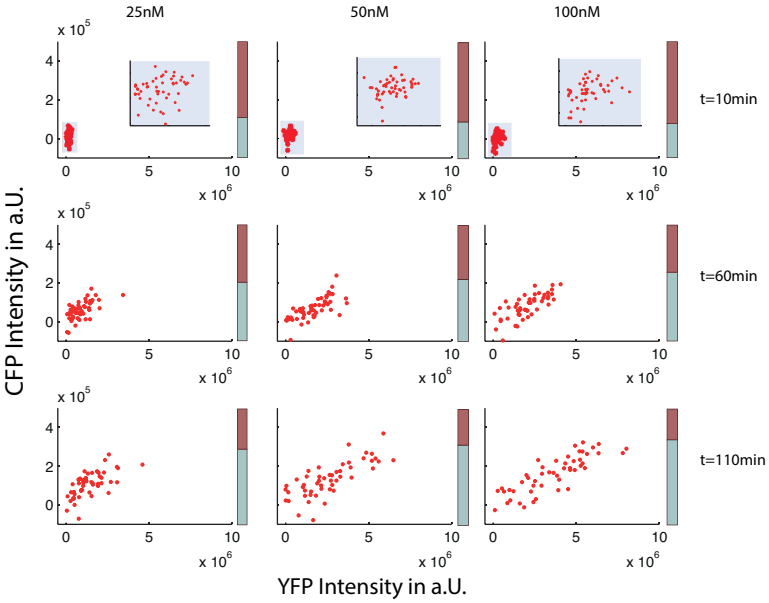

Figure 22: Intrinsic and extrinsic noise revealed by microscopy in a strain that contains both CFP and YFP reporters driven by the pGAL1-promoter. Expression was induced by addition of β-estradiol at different concentrations. Rough gates were applied time-point-wise on the CFP and YFP channels in order to remove dead cells and segmentation errors. The remaining cells (e.g., around 150-300) were used for estimating intrinsic and extrinsic noise. 50 randomly selected cells are plotted. Data at $t = 10\,\text{min}$ is also shown on a zoomed scale for better illustration (inset scatter plots). The bar on each graph represents the percentage of intrinsic (dark red) and extrinsic (green) noise over total noise.

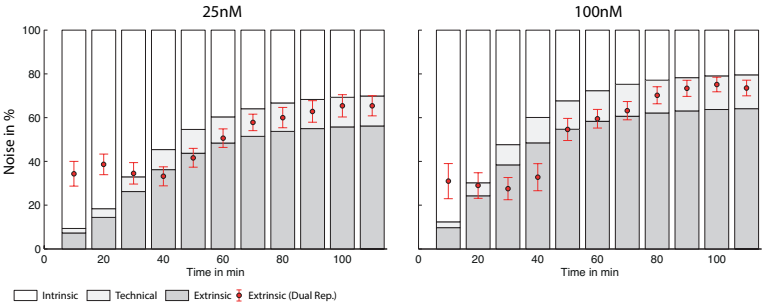

Figure 23: Comparison between predicted and experimentally determined noise contributions for a 25 nM and 100 nM pulse of β-estradiol. The analysis was performed such as described in the main text. Whiskers indicate standard errors of the experimentally obtained quantities.

SUPPLEMENTARY NOTE 1:

THEORY AND ALGORITHMS

MARGINAL PROCESS DYNAMICS

Let $X \mid (S, Z)$ be a conditional CTMC describing the time evolution of a reaction network with ν reaction channels and associated kinetic parameters $C = \{C_1, \ldots, C_\nu\}$. The dynamics are governed by a set of fixed (i.e., shared) intrinsic parameters $S = \{S_i \mid i = 1, \ldots, I\}$ as well as a set of extrinsic factors $Z = \{Z_i \mid i = 1, \ldots, J\}$ that randomly vary between cells. Although a more general treatment is possible, we restrict ourselves to the case where both the intrinsic parameters and extrinsic factors are kinetic rate constants such that $Z \cup S = C$. For a convenient notation and without loss of generality, we assume a particular ordering of the kinetic parameters, i.e., $C = \{Z_1, \ldots, Z_J, S_1, \ldots, S_I\}$. Then, with $X(t) = x$ as the state of the CTMC at time t, the propensities for the next reaction are given by $h_j(x, z_j) = z_j g_j(x)$ for $j = 1, \ldots, J$ and $h_i(x, s_i) = s_i g_i(x)$ for $i = J+1, \ldots, \nu$.

Our goal is to find a dynamic description of the marginal process $X \mid (S, A)$, where the extrinsic parameters are integrated out, such that their randomness is directly encoded in the resulting process dynamics. Within the theory of counting processes, such studies are centered around the *innovation theorem* [1]. In the context of CTMCs, a similar problem has been investigated in [2], where the author studied the marginal dynamics of a simple three-state Markov Chain with random intensities. In the following we prove that the construction of a marginal process is possible for *arbitrary* reaction networks with propensity functions linear in Z (e.g., such as for mass-action kinetics).

Proposition 1 *The propensities of the reactions with index* $j = 1, \ldots, J$ *of the marginal process* $X \mid (S, A)$ *are given by*

$$h_j(\mathbf{x}, t) = \mathbb{E}\left[Z_j \mid \mathbf{x}, a\right] g_j(x(t)), \tag{39}$$

with $\mathbb{E}\left[Z_j \mid \mathbf{x}, a\right]$ *as the conditional expectation of* Z_j *given a sample path* $\mathbf{x} = \{x(s) \mid s \in [0, t]\}$. *All other propensities, i.e., those corresponding to reaction indices* $i = J+1, \ldots, \nu$ *remain unchanged.*

Proof 3 *Let* $P\left(X(t+dt) = x(t) + \Delta_j \mid z_j, x(t)\right)$ *denote the probability that reaction* j *fires within the interval* $[t, t+dt)$ *given the current state* $x(t)$, *where*

Δ_j corresponds to the stoichiometric change vector of reaction j. Then, for the marginal jump probability we obtain

$$\begin{aligned}
&P\left(X(t+dt) = x(t) + \Delta_j \mid \mathbf{x}, a\right) \\
&\quad = \int_{\mathcal{Z}^J} P\left(X(t+dt) = x(t) + \Delta_j, z \mid \mathbf{x}, a\right) dz \\
&\quad = \int_{\mathcal{Z}^J} P\left(X(t+dt) = x(t) + \Delta_j \mid x(t), z_j\right) p(z \mid \mathbf{x}, a) dz \\
&\quad = \left(\int_{\mathcal{Z}^J} z_j p(z \mid \mathbf{x}, a) dz\right) g_j(x(t)) dt \\
&\quad = \mathbb{E}\left[Z_j \mid \mathbf{x}, a\right] g_j(x(t)) dt \\
&\quad = h_j(\mathbf{x}, t) dt.
\end{aligned} \tag{40}$$

■

Remark 2 (Moment Generating Function (MGF) Representation) *We know from [245, 124] that the likelihood function of a path $\mathbf{x}$ with respect to the parameters z is given by*

$$p\left(\mathbf{x} \mid z\right) \propto \prod_{i=1}^{J} z_i^{r_i} \exp\left\{-\sum_{i=1}^{J}\left(z_i \int_0^t g_i(x(\tau)) d\tau\right)\right\}, \tag{41}$$

where r_i counts the number of reactions of type i.

Then, by applying Bayes' formula, the conditional expectation in (39) can be written as

$$\begin{aligned}
\mathbb{E}\left[Z_j \mid \mathbf{x}, a\right] &= \int_{\mathcal{Z}} z_j p(z_j \mid \mathbf{x}, a) dz_j \\
&= \int_{\mathcal{Z}} z_j \frac{p(\mathbf{x} \mid z_j) p(z_j \mid a)}{p(\mathbf{x} \mid a)} dz_j \\
&= \int_{\mathcal{Z}^J} z_j \frac{p(\mathbf{x} \mid z) p(z \mid a)}{p(\mathbf{x} \mid a)} dz.
\end{aligned} \tag{42}$$

Hence, the marginal reaction hazard can be reformulated as

$$\begin{aligned}
h_j(\mathbf{x}, t) &= \mathbb{E}\left[Z_j \mid \mathbf{x}, a\right] g_j(x(t)) \\
&= \left(\int_{\mathcal{Z}^J} z_j \frac{p(\mathbf{x} \mid z) p(z \mid a)}{p(\mathbf{x} \mid a)} dz\right) g_j(x(t)) \\
&= \left(\int_{\mathcal{Z}^J} z_j \frac{p(\mathbf{x} \mid z) p(z \mid a)}{\int_{\mathcal{Z}^J} p(\mathbf{x} \mid z) p(z \mid a) dz} dz\right) g_j(x(t)) \\
&= \frac{\mathbb{E}\left[z_j p\left(\mathbf{x} \mid z\right) \mid a\right]}{\mathbb{E}\left[p\left(\mathbf{x} \mid z\right) \mid a\right]} g_j(x(t)) \\
&= \frac{\mathbb{E}\left[z_j \prod_{i=1}^{J} z_i^{r_i} \exp\left\{-\sum_{i=1}^{J}\left(z_i \int_0^t g_i(x(\tau)) d\tau\right)\right\} \mid a\right]}{\mathbb{E}\left[\prod_{i=1}^{J} z_i^{r_i} \exp\left\{-\sum_{i=1}^{J}\left(z_i \int_0^t g_i(x(\tau)) d\tau\right)\right\} \mid a\right]} g_j(x(t)).
\end{aligned} \tag{43}$$

We note that the expectations in the numerator and denominator can be rewritten as higher-order partial derivatives of the MGF[1] of $Z \mid (A = a)$ *[2] and we arrive at*

$$h_j(\mathbf{x}, t) = \left[\frac{\partial \prod_{i=1}^{J} \partial^{r_i}}{\partial \sigma_j \prod_{i=1}^{J} \partial \sigma_i^{r_i}} G_{Z|a}(\sigma_1, \ldots, \sigma_J) \left(\frac{\prod_{i=1}^{J} \partial^{r_i}}{\prod_{i=1}^{J} \partial \sigma_i^{r_i}} G_{Z|a}(\sigma_1, \ldots, \sigma_J) \right)^{-1} \right] g_j(x(t)), \tag{44}$$

with $G_{Z|a}$ *as the MGF of* $Z \mid (A = a)$ *and* $\sigma_i \equiv -\int_0^t g_i(x(\tau))d\tau$.

Example 1 (Univariate Gamma Distribution) *We assume a one-dimensional Gamma-distributed extrinsic parameter* $Z \mid (A = a) \sim \mathcal{G}(\alpha, \beta)$ *with* $a = \{\alpha, \beta\}$ *and reaction index* 1. *The MGF is known to be*

$$G_{Z|a}(\sigma) = \frac{\beta^\alpha}{(\beta - \sigma)^\alpha}$$

and the i*-th derivative becomes*

$$\frac{d^i}{d\sigma^i} G_{Z|a}(\sigma) = \frac{\Gamma(\alpha + i)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta - \sigma)^{\alpha+i}}.$$

Hence, by substituting into (44) the marginal hazard function is given by

$$h_1(\mathbf{x}, t) = \frac{\alpha + r_1}{\beta + \int_0^t g_1(x(\tau))d\tau} g_1(x(t)). \tag{45}$$

Example 2 (Conditioning on Covariates) *Often, covariates* V *of* Z *can be obtained experimentally (e.g., morphological features). We assume knowledge of a measurement density such that* $V \mid (Z = z, B = b) \sim p(v \mid z, b)$. *In this case, the marginal hazard functions become*

$$h_j(\mathbf{x}, v, t) = \left[\frac{\partial \prod_{i=1}^{J} \partial^{r_i}}{\partial \sigma_j \prod_{i=1}^{J} \partial \sigma_i^{r_i}} G_{Z|v,a,b}(\sigma_1, \ldots, \sigma_J) \left(\frac{\prod_{i=1}^{J} \partial^{r_i}}{\prod_{i=1}^{J} \partial \sigma_i^{r_i}} G_{Z|v,a,b}(\sigma_1, \ldots, \sigma_J) \right)^{-1} \right] g_j(x(t)),$$

where $G_{Z|v,a,b}$ is the MGF of $Z \mid (V = v, A = a, B = b) \sim p(z \mid v, a, b) \propto p(v \mid z, b)p(z \mid a)$. In case of Example 1, i.e., $Z \mid (A = a) \sim \mathcal{G}(\alpha, \beta)$ and $V \mid (Z = z, B = b) \sim \mathcal{G}(\rho, \phi z)$ with $b = \{\rho, \phi\}$, we obtain

$$\begin{aligned} p(z \mid v, a, b) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp\{-\beta z\} \frac{(\phi z)^\rho}{\Gamma(\rho)} v^{\rho-1} \exp\{-\phi z v\} \\ &= \frac{\beta^\alpha \phi^\rho v^{\rho-1}}{\Gamma(\alpha)\Gamma(\rho)} z^{\alpha+\rho-1} \exp\{-z(\beta + \phi v)\}, \end{aligned}$$

1 The moment generating function of a random vector $Z = (Z_1, \ldots, Z_n)$ is defined as $\mathbb{E}\left[e^{\sigma_1 Z_1 + \ldots + \sigma_n Z_n}\right]$.

and hence, $Z \mid (V = v, A = a, B = b) \sim \mathcal{G}(\alpha + \rho, \beta + \phi v)$. Then, the marginal reaction hazard is given by

$$h_1(\mathbf{x}, v, t) = \frac{\alpha + \rho + r_1}{\beta + \phi v + \int_0^t g_1(x(\tau))d\tau} g_1(x(t)). \tag{46}$$

Since a more detailed analysis of the marginal stochastic process is beyond the scope of the present study, the above derivations are restricted to parts that are needed for the inference scheme and the performed case studies. However, we remark that such a marginalization can also be performed if the distributional assumptions on Z are different than Gamma [3, 255].

Stochastic Simulation of Marginal Dynamics

As indicated in the main text, the marginal process is governed by time-dependent hazard functions. However, exact simulation can be performed using standard methods such as the first reaction method [76]. In contrast to standard SSA, waiting times Δt_j are computed for each reaction individually by solving

$$\int_0^{\Delta t_j} h_j(\mathbf{x}, t)dt = -\ln \tau_j, \tag{47}$$

with $h_j(\mathbf{x}, t)$ as the reaction hazard for reaction j and τ_j as a random number drawn from $\mathcal{U}(0, 1)$. Subsequently, the next reaction index i is selected according to the smallest waiting time Δt_i. For homogeneous reactions we obtain $\Delta t_j = -(c_j g_j(x))^{-1} \ln \tau_j$ with x as the current state of the system. For instance, in presence of Gamma-type heterogeneity, the solution of (47) becomes

$$\Delta t_j = -\frac{G_j(\mathbf{x}, t) + \beta - \exp\left[-\frac{\ln \tau_j}{\alpha + r_j} + \ln\left(G_j(\mathbf{x}, t) + \beta\right)\right]}{g_j(x)} \tag{48}$$

with $G_j(\mathbf{x}, t) = \int_0^t g_j(x(s))ds$.

In the following, we will frequently consider cases, where an ensemble of M trajectories (each of them corresponding to particular cell) need to be simulated. Moreover, sample paths are not simulated on the full measurement interval, but rather established in a sequential manner (i.e., successively extended as time increases). We know from above that the marginal process can be simulated at any time conditional on its history. More specifically, each propensity function depends on certain path statistics, summarizing the past of the corresponding reaction channel. In case of the reactions that correspond to the shared kinetic parameters, the summary statistics contain information about a particular cell, as well as all other cells in the population. As a consequence, the marginal trajectories become explicitly dependent (as opposed to the original Markovian model). In contrast, the reaction propensities corresponding to the extrinsic factors only depend on their own history due to the assumption that

each cell's extrinsic factors are independently drawn. In the following, we denote by $T_l^k = (T_1[k,l], \ldots, T_J[k,l])$ with $T_j[k,l] = (R_j[k,l], G_j[k,l])$ the collection of path statistics required for simulating the k-th cell at time t_l. The statistics corresponding to the intrinsic reactions $i = J+1, \ldots, \nu$ are inductively defined as

$$\begin{aligned} k = 2, \ldots, M: \qquad & R_i[k,l] = R_i[k-1,l] + r_i(\mathbf{x}_{l:l+1}^{k-1}) \\ & G_i[k,l] = G_i[k-1,l] + \int_{t_l}^{t_{l+1}} g_i(x^{k-1}(\tau))d\tau \\ k = 1: \qquad & R_i[1,l] = R_i[M,l-1] + r_i(\mathbf{x}_{l-1:l}^{M}) \\ & G_i[1,l] = G_i[M,l-1] + \int_{t_{l-1}}^{t_l} g_i(x^{M}(\tau))d\tau \end{aligned} \tag{49}$$

with $R_i[1,1] = G_i[1,1] = 0$ and $r_i(\mathbf{x})$ as the number of reactions of type i in $\mathbf{x}$. Similarly, we obtain for the statistics corresponding to the extrinsic reactions $j = 1, \ldots, J$

$$\begin{aligned} R_j[k,l] &= R_j[k,l-1] + r_j(\mathbf{x}_{l-1:l}^{k}) \\ G_j[k,l] &= G_j[k,l-1] + \int_{t_{l-1}}^{t_l} g_j(x^{k}(\tau))d\tau \end{aligned} \tag{50}$$

with $R_j[k,1] = G_j[k,1] = 0$ for $k = 1, \ldots, M$.

For instance, consider M trajectories on $N-1$ subintervals. For a set of hyperparameters a, the joint path-likelihood can be compactly written as

$$p(\mathbf{x}_{1:N}^1, \ldots, \mathbf{x}_{1:N}^M \mid a) = \prod_{l=2}^{N} \prod_{m=1}^{M} p(\mathbf{x}_{l-1:l}^m \mid x_{l-1}^m, T_{l-1}^m, a). \tag{51}$$

SEQUENTIAL MARKOV CHAIN MONTE CARLO (SMCMC)

Several techniques for the inference of partially observed stochastic reaction systems have been recently proposed [8, 258, 84, 83, 171]. Generally, such methods can be divided into two subgroups, i.e., analytical [258, 171] and sampling-based approaches [8, 84, 83]. While the former are beneficial in terms of computational complexity, they typically rely on approximations of the target posterior distribution. In contrast, sampling-based approaches are capable of drawing samples from the exact posterior distribution, but on the downside, require a larger computational effort, especially when dealing with a high-dimensional parameter- and state-space. Here, we combine a sampling-based approach with the analytical marginalization described above. Typically, such marginalized inference schemes can profit from expedient statistical properties such as a reduced variance of the resulting parameter- and state estimates. In the following, we give a detailed description of the proposed method.

As described in the Online Methods, the hierarchical state space model from Figure 1B can be marginalized with respect to the kinetic parameters

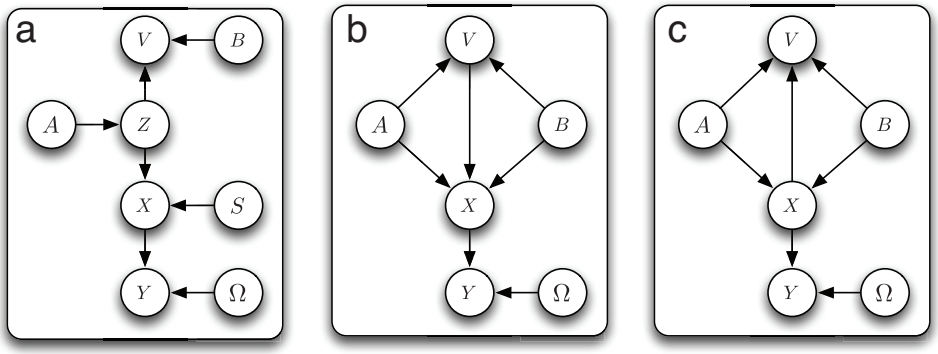

Figure 24: Graphical explanation of the marginalization. (a) Original Bayesian network. (b, c) Marginalized Bayesian networks. Both models represent valid Bayesian networks for the marginalized model (i.e., are mathematically equivalent), whereas they differ in the causality between V and X. Note that for clarity, individual time points are not represented separately in the above illustration.

as well as the extrinsic factors such that DPP can be applied. However, the resulting model quantities, i.e., A, B and Ω remain in the model. This is illustrated in Figure 24a, Figure 24b and Figure 24c (see caption for further details).

The joint distribution of the marginalized model is given by

$$\begin{aligned} &p(a, b, \omega, \mathbf{x}^1_{1:N}, \ldots, \mathbf{x}^M_{1:N}, y^1_{1:N}, \ldots, y^M_{1:N}, v^1, \ldots, v^M) = \\ &\quad \left[\prod_{m=1}^{M} \left(\prod_{l=1}^{N} p(y^m_l \mid x^m_l, \omega)\right) p(v^m \mid \mathbf{x}^m_{1:N}, b, a)\right] \\ &\quad p(\mathbf{x}^1_{1:N}, \ldots, \mathbf{x}^M_{1:N} \mid a, b) p(a) p(b) p(\omega). \end{aligned} \tag{52}$$

We are interested in sampling from the posterior distribution

$$\begin{aligned} &p(\mathbf{x}^1_{1:N}, \ldots, \mathbf{x}^M_{1:N}, a, b, \omega \mid \{y^1_{1:N}, v^1\}, \ldots, \{y^M_{1:N}, v^M\}) \\ &\quad \propto \left[\prod_{m=1}^{M} \left(\prod_{l=1}^{N} p(y^m_l \mid x^m_l, \omega)\right) p(v^m \mid \mathbf{x}^m_{1:N}, b, a)\right] \\ &\qquad p(\mathbf{x}^1_{1:N}, \ldots, \mathbf{x}^M_{1:N} \mid a, b) p(a) p(b) p(\omega) \\ &\quad = \left[\prod_{m=1}^{M} \left(\prod_{l=1}^{N} p(y^m_l \mid x^m_l, \omega)\right) p(v^m \mid b, a)\right] \\ &\qquad p(\mathbf{x}^1_{1:N}, \ldots, \mathbf{x}^M_{1:N} \mid a, b, v^1, \ldots, v^M) p(a) p(b) p(\omega), \end{aligned} \tag{53}$$

where $\{\mathbf{x}^1_{1:N}, \ldots, \mathbf{x}^M_{1:N}, a, b, \omega\}$ denote the unknown (i.e. *latent*) quantities. As sampling from the full latent space is practically impossible, we stick to

a recursive Bayesian inference procedure, where the posterior distribution at time t_l is computed from the posterior distribution at time t_{l-1} as

$$\begin{aligned} &p\left(\mathbf{x}_{1:l}^1, \ldots, \mathbf{x}_{1:l}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \ldots, \{y_{1:l}^M, v^M\}\right) \\ &\quad \propto \left[\prod_{m=1}^{M} p(y_l^m \mid x_l^m, \omega) p(\mathbf{x}_{l-1:l}^m \mid x_{l-1}^m, T_{l-1}^m, a, b, v^m)\right] \\ &\quad \times p\left(\mathbf{x}_{1:l-1}^1, \ldots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l-1}^1, v^1\}, \ldots, \{y_{1:l-1}^M, v^M\}\right), \end{aligned} \tag{54}$$

By exploiting the recursive relation of the posterior distribution, the original inference problem breaks up into a sequence of smaller problems which are easier solve. The resulting algorithms go under the name of *sequential Monte Carlo* (SMC) methods [54, 53]. When applied to combined parameter- and state inference problems, standard SMC methods are likely to suffer from particle degeneracy, since the static parameters cannot take values different from their initialization at time t_1. One strategy to overcome such problems is to randomly perturb the static parameters at each time step, in order to maintain diversity among the particles [219]. Practically, this means that the parameter values of a drawn particle are discarded and newly sampled. Importantly, the resampling should be performed such that the new parameter values are again a valid sample from the posterior distribution. This is generally achieved be applying an *invariant kernel* to the original parameter values.

For that sake, note that the posterior can be written as

$$\begin{aligned} &p\left(\mathbf{x}_{1:l}^1, \ldots, \mathbf{x}_{1:l}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \ldots, \{y_{1:l}^M, v^M\}\right) \\ &\quad = p\left(a, b, \omega \mid \{\mathbf{x}_{1:l}^1, y_{1:l}^1, v^1\}, \ldots, \{\mathbf{x}_{1:l}^M, y_{1:l}^M, v^M\}\right) \\ &\quad \times p\left(\mathbf{x}_{1:l}^1, \ldots, \mathbf{x}_{1:l}^M \mid \{y_{1:l}^1, v^1\}, \ldots, \{y_{1:l}^M, v^M\}\right). \end{aligned} \tag{55}$$

Hence, diversified samples $\{\tilde{\mathbf{x}}_{1:l}^1, \ldots, \tilde{\mathbf{x}}_{1:l}^M, \tilde{a}, \tilde{b}, \tilde{\omega}\}$ can be drawn by first sampling trajectories

$$\left\{\tilde{\mathbf{x}}_{1:l}^1, \ldots, \tilde{\mathbf{x}}_{1:l}^M\right\} \sim p\left(\mathbf{x}_{1:l}^1, \ldots, \mathbf{x}_{1:l}^M \mid \{y_{1:l}^1, v^1\}, \ldots, \{y_{1:l}^M, v^M\}\right) \tag{56}$$

and subsequently drawing

$$\left\{\tilde{a}, \tilde{b}, \tilde{\omega}\right\} \sim p\left(a, b, \omega \mid \{\tilde{\mathbf{x}}_{1:l}^1, y_{1:l}^1, v^1\}, \ldots, \{\tilde{\mathbf{x}}_{1:l}^M, y_{1:l}^M, v^M\}\right). \tag{57}$$

Within an SMC framework it is straight-forward to sample from the marginal distribution (56), given that the full posterior at a certain time step is available as a set of (possibly weighted) particles. Once a sample is drawn, marginalization is carried out by only considering the variables of interest (i.e., $\{\tilde{\mathbf{x}}_{1:l}^1, \ldots, \tilde{\mathbf{x}}_{1:l}^M\}$). However, it appears to be complicated to directly sample from (57) using standard techniques such as the Metropolis-Hastings (M-H) algorithm. More specifically, such sampling methods rely on appropriate proposal distributions that are typical hard to find - especially if the sampling space is large. In such cases it can be beneficial to use a Gibbs-like Metropolis-Hastings sampler [152], where subsets of variables

are resampled conditionally on the other variables, i.e., are drawn from the respective *full conditional* distributions[2]. Those distributions satisfy the invariance requirement [188] and can be easily identified from the underlying Bayesian network. On the one hand, some of the full conditional distributions might be of standard form, such that samples can be drawn straight away (e.g., if the distribution is Gaussian). On the other hand, it is often less challenging to find good proposal distributions for a single quantity such that better acceptance ratios can be achieved. Assume that the set of all variables in a network is partitioned into subsets $\mathcal{U} = \{\mathcal{U}_1, \ldots, \mathcal{U}_J\}$. Then the subset $\mathcal{U}_j$ is independent of all other variables in the Bayesian network when conditioned on its Markov blanket $\mathcal{MB}(\mathcal{U}_j)$, defined as the set of parents, children and the children's parents of $\mathcal{U}_j$ [119]. Hence, the full conditional distribution over $\mathcal{U}_j$ is given by $p(\mathcal{U}_j \mid \overline{\mathcal{U}_j}) = p(\mathcal{U}_j \mid \mathcal{MB}(\mathcal{U}_j))$. Defining subsets $\mathcal{U}_1 = \{\Omega\}$ and $\mathcal{U}_2 = \{A, B\}$ we obtain

$$\tilde{\omega} \sim p\left(\omega \mid \left\{\tilde{x}_1^1, y_1^1\right\}, \ldots, \left\{\tilde{x}_l^M, y_l^M\right\}\right) \tag{58}$$

$$\left\{\tilde{a}, \tilde{b}\right\} \sim p\left(a, b \mid \left\{\tilde{\mathbf{x}}_{1:l}^1, v^1\right\}, \ldots, \left\{\tilde{\mathbf{x}}_{1:l}^M, v^M\right\}\right). \tag{59}$$

In the following, we will discuss in detail how to resample the individual quantities in (56), (58) and (59).

$$\begin{aligned} &p\left(\mathbf{x}_{1:l}^1, \mathbf{x}_{1:l-1}^2, \ldots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l}^1, v^1\}, \ldots, \{y_{1:l-1}^M, v^M\}\right) \\ &\quad \propto p(y_l^1 \mid x_l^1, \omega) p(\mathbf{x}_{l-1:l}^1 \mid x_{l-1}^1, a, b, v^1, T_{l-1}^1) \\ &\quad \times p\left(\mathbf{x}_{1:l-1}^1, \ldots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l-1}^1, v^1\}, \ldots, \{y_{1:l-1}^M, v^M\}\right). \end{aligned} \tag{60}$$

Note that we can easily sample from (60) using a M-H criterion. If we propose samples

$$\begin{aligned} &\{\tilde{\mathbf{x}}_{1:l}^1, \tilde{\mathbf{x}}_{1:l-1}^2, \ldots, \tilde{\mathbf{x}}_{1:l-1}^M, \tilde{a}, \tilde{b}, \tilde{\omega}\} \\ &\quad \sim p(\mathbf{x}_{l-1:l}^1 \mid x_{l-1}^1, a, b, v^1, T_{l-1}^1) \\ &\quad \times p\left(\mathbf{x}_{1:l-1}^1, \ldots, \mathbf{x}_{1:l-1}^M, a, b, \omega \mid \{y_{1:l-1}^1, v^1\}, \ldots, \{y_{1:l-1}^M, v^M\}\right), \end{aligned} \tag{61}$$

the acceptance probability reduces to

$$\gamma_x^1 = \min\left\{1, \frac{p(y_l^1 \mid \tilde{x}_l^1, \tilde{\omega})}{p(y_l^1 \mid x_l^1, \omega)}\right\}. \tag{62}$$

We proceed analogously for the remaining cells until we have obtained the full posterior distribution over all cells at time t_l. Note that samples from $p(\mathbf{x}_{l-1:l}^m \mid x_{l-1}^m, a, b, v^m, T_{l-1}^m)$ are easily obtained using the stochastic simulation algorithm from section 6. For all algorithms considered in this work, we adopt this choice of proposal distribution.

2 Such a scheme can be understood as component-wise modification of the standard M-H sampler.

Resampling the Measurement Parameters

In most realistic cases, statistics of the measurement noise are unknown and hence, need to be included into the inference. We note that the measurement parameters Ω are independent of all other nodes in the network given the state at the sampling points x_l^m and measurements y_l^m for all $m = 1, \ldots, M$, i.e., conditional on the set $\mathcal{MB}(\Omega) = \{\{X_i^m, Y_i^m\} \mid i = 1, \ldots, l \wedge m = 1, \ldots, M\}$. Therefore, the full conditional can be written as

$$p(\omega \mid \{X_1^1, Y_1^1\}, \ldots, \{X_l^M, Y_l^M\}) \propto \left(\prod_{m=1}^{M} \prod_{i=1}^{l} p(y_i^m \mid x_i^m, \omega) \right) p(\omega). \quad (63)$$

Depending on the specific structure of measurement likelihood function $p(y_i^m \mid x_l^m, \omega)$, the corresponding unknown parameters Ω and the prior $p(\omega)$, eq. (63) might be of standard form. For instance, for a normally or log-normally distributed measurement noise with unknown scaling parameter $\Omega := \sigma$, gamma priors $\Omega \sim \mathcal{G}(\alpha_\omega, \beta_\omega)$ can be used. In particular, samples from the resulting full conditional are obtained as $\tilde{\omega} = \sqrt{1/\tilde{\tau}}$ with

$$\tilde{\tau} \sim \mathcal{G}\left(\frac{lM}{2} + \alpha_\omega, \frac{\sum_{m=1}^{M} \sum_{i=1}^{l} (x_i^m - y_i^m)^2}{2} + \beta_\omega \right) \quad (64)$$

in case of a normal measurement noise and with

$$\tilde{\tau} \sim \mathcal{G}\left(\frac{lM}{2} + \alpha_\omega, \frac{\sum_{m=1}^{M} \sum_{i=1}^{l} (\ln x_i^m - \ln y_i^m)^2}{2} + \beta_\omega \right) \quad (65)$$

in case of a log-normal measurement noise.

Resampling the Extrinsic Statistics and Morphological Shape Parameters

We know from (59) that A and B can be resampled conditional on $\left\{ \left\{ \mathbf{X}_{1:l}^1, V^1 \right\}, \ldots, \left\{ \mathbf{X}_{1:l}^M, V^M \right\} \right\}$. The corresponding full conditional distribution takes the form

$$\begin{aligned} &p\left(a, b \mid \{\tilde{\mathbf{x}}_{1:l}^1, v^1\}, \ldots, \{\tilde{\mathbf{x}}_{1:l}^M, v^M\}\right) = \\ &\quad \frac{1}{Z} \left(\prod_{m=1}^{M} p(\tilde{\mathbf{x}}^m \mid v^m, a, b) p(v^m \mid a, b) \right) p(a) p(b). \end{aligned} \quad (66)$$

Evaluation of eq. (66) requires knowledge of the marginal path likelihood functions $p(\mathbf{x} \mid v, a, b^m)$. We assume the same configuration as in Example 2, i.e., one-dimensional extrinsic factors Z and corresponding co-

variates V. The path likelihood given the shared parameters S and extrinsic factors Z is given by

$$\begin{aligned} p(\mathbf{x} \mid s, z) &\propto \left(\prod_{i=J+1}^{\nu} s_i^{r_i} \exp\left\{ -s_i \int_0^t g_i(x(\tau))d\tau \right\} \right) \times z^{r_1} \\ &\qquad \exp\left\{ -z \int_0^t g_1(x(\tau))d\tau \right\} \\ &= \prod_{i=J+1}^{\nu} f_i(\mathbf{x}, s_i) \times f_1(\mathbf{x}, z). \end{aligned} \tag{67}$$

Due to the product form of (67), only the function f_1 corresponding to the extrinsic factor Z will depend on A and B, i.e., the marginalized likelihood function becomes

$$\begin{aligned} f_1(\mathbf{x}, a, b, v) &= \int_{\mathcal{Z}} f_1(\mathbf{x}, z) p(z \mid a) p(v \mid z, b) dz \\ &\propto \int_{\mathcal{Z}} f_1(\mathbf{x}, z) p(z \mid v, a, b) dz \\ &= \mathbb{E}_{\cdot}\left[f_1(\mathbf{x}, z) \mid v, a, b \right] \end{aligned} \tag{68}$$

In case of Gamma-distributed random effects $Z \mid (A = a) \sim \mathcal{G}(\alpha, \beta)$ and covariates $V \mid (Z = z, B = b) \sim \mathcal{G}(\rho, \phi z)$, we know that $Z \mid (V = v, A = a, B = b) \sim \mathcal{G}(\alpha + \rho, \beta + \phi v)$. Hence, we further obtain

$$\begin{aligned} f_1(\mathbf{x}, a, b, v) &= \frac{(\beta + \phi v)^{\alpha+\rho}}{\Gamma(\alpha + \rho)} \int_{\mathcal{Z}} z^{(r_1 + \alpha + \rho - 1)} \\ &\qquad \exp\left\{ -z \left[\beta + \phi v + \int_0^t g_1(x(s))ds \right] \right\} dz \\ &= \frac{(\beta + \phi v)^{\alpha+\rho} \Gamma(\alpha + \rho + r_1)}{\Gamma(\alpha + \rho)} \\ &\qquad \left[\beta + \phi v + \int_0^t g_1(x(s))ds \right]^{-(\alpha+\rho+r_1)}. \end{aligned} \tag{69}$$

Similarly, the density $p(v \mid a, b)$ is found to be

$$\begin{aligned} p(v \mid a, b) &= \int_{\mathcal{Z}} p(v \mid z, b) p(z \mid a) dz \\ &= \frac{\beta^{\alpha} \phi^{\rho} v^{\rho-1} \Gamma(\alpha + \rho)(\beta + \phi v)^{-(\alpha+\rho)}}{\Gamma(\alpha)\Gamma(\rho)}. \end{aligned} \tag{70}$$

Finally, we can rewrite (66) as

$$\begin{aligned} &p\left(a, b \mid \{\tilde{\mathbf{x}}_{1:l}^1, v^1\}, \ldots, \{\tilde{\mathbf{x}}_{1:l}^M, v^M\} \right) \\ &\qquad = \frac{1}{\widehat{Z}} \left(\prod_{m=1}^{M} f_1(\mathbf{x}^m, a, b, v^m) p(v^m \mid a, b) \right) p(a) p(b). \end{aligned} \tag{71}$$

Due to the complicated structure of (69) and (70), we cannot directly sample from the full conditional such that we again make use of a M-H step, where proposed samples $\{\tilde{a}, \tilde{b}\} \sim q(\cdot, \cdot)$ are accepted with probability

$$\gamma_{A,B} = \min\left\{1, \frac{\prod_{m=1}^{M} f_1(\mathbf{x}^m, \tilde{a}, \tilde{b}, \nu^m) p(\nu^m \mid \tilde{a}, \tilde{b}) p(\tilde{a}) p(\tilde{b}) q(a, b)}{\prod_{m=1}^{M} f_1(\mathbf{x}^m, a, b, \nu^m) p(\nu^m \mid a, b) p(a) p(b) q(\tilde{a}, \tilde{b})}\right\}. \quad (72)$$

For the case studies considered here, we chose q to be a multivariate log-normal distribution.

Full Posterior Reconstruction

Since we have marginalized the joint distribution with respect to certain variables, execution of the inference scheme can only deliver marginal posterior distributions. More specifically, the algorithm returns a set of P particles consisting of the variables $\{\mathbf{X}^1_{1:l-1}, \ldots, \mathbf{X}^M_{1:l-1}, A, B, \Omega\}$, while the shared and extrinsic parameters S and $Z^1, \ldots, Z^M$ are not included. However, the particle distribution over all variables can be easily reconstructed via the law of conditional probability, i.e.,

$$\begin{aligned} &p(s, z^1, \ldots, z^M, \mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l}, a, b, \omega \mid \{y^1_{1:l}, \nu^1\}, \ldots, \{y^M_{1:l}, \nu^M\}) \\ &\quad = p\left(s \mid \mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^1_{1:l}\right) \left[\prod_{m=1}^{M} p(z^m \mid \mathbf{x}^m_{1:l}, \nu^m, b)\right] \\ &\qquad \times p(\mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l}, a, b, \omega \mid \{y^1_{1:l}, \nu^1\}, \ldots, \{y^M_{1:l}, \nu^M\}). \end{aligned} \quad (73)$$

This implies that a particle from the full posterior distribution can be constructed by first drawing a particle from the marginal distribution and subsequently sampling S and $Z^1, \ldots, Z^M$ conditional on that particle. If we assume Gamma-type prior distributions for each of the kinetic parameters, i.e., $S_i \sim \mathcal{G}(\kappa_i, \chi_i)$, the corresponding conditional distribution $p(s_i \mid \mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l})$ is again Gamma, i.e., $\mathcal{G}(\kappa_i + R_i, \chi_i + G_i)$ with $R_i = \sum_{m=1}^{M} r_i^m$ and $G_i = \sum_{m=1}^{M} \int_0^{t_l} g_i(x^m(\tau)) d\tau$.

Moreover, the marginal parameter posterior can be written as a multivariate compound Gamma distribution

$$p(s \mid \{y^1_{1:l}, \nu^1\}, \ldots, \{y^M_{1:l}, \nu^M\}) = \mathbb{E}\left[p(s \mid \mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l})\right] \quad (74)$$

with $p(s \mid \mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l})$ given by the product of the individual conditional distributions, i.e.,
$\prod_{i=J+1}^{\nu} \mathcal{G}(\kappa_i + R_i, \chi_i + G_i)$. Note that R_i and G_i are functions of the sample paths and that the expectation in (74) is taken with respect to the smoothing distribution

$$p(\mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l} \mid \{y^1_{1:l}, \nu^1\}, \ldots, \{y^M_{1:l}, \nu^M\}) \approx \frac{1}{P} \sum_{p=1}^{P} \mathbb{1}_{\Lambda_S^{(p)}}\left(\mathbf{x}^1_{1:l}, \ldots, \mathbf{x}^M_{1:l}\right), \quad (75)$$

with $\Lambda_S^{(p)} = \{\mathbf{x}_{1:l}^1, \ldots, \mathbf{x}_{1:l}^M\}^{(p)}$ collecting the M sample paths of the p-th particle. Consequently - within the finite particle representation - (74) is approximated as a sum of Gamma distributions.

Analogously, we can perform the reconstruction of the marginal (and joint) posterior for the extrinsic factor Z, which for clarity is again assumed to be one-dimensional. Note however, that according to the product form in (73), this can be carried out independently for each cell. For the m-th cell we compute the compound density

$$p(z^m \mid \{y_{1:l}^1, \nu^1\}, \ldots, \{y_{1:l}^M, \nu^M\}) = \mathbb{E}\left[p(z^m \mid \mathbf{x}_{1:l}^m, \nu^m, a, b)\right], \tag{76}$$

where the expectation is taken with respect to the distribution

$$p(\mathbf{x}_{1:l}^m, a, b \mid \{y_{1:l}^1, \nu^1\}, \ldots, \{y_{1:l}^M, \nu^M\}) \approx \frac{1}{P} \sum_{p=1}^{P} \mathbb{1}_{\Lambda_Z^{(p)}}(\mathbf{x}_{1:l}^m, a, b), \tag{77}$$

with $\Lambda_Z^{(p)} = \{\mathbf{x}_{1:l}^m, a, b\}^{(p)}$. The full conditional distribution inside the expectation factorizes as

$$p(z^m \mid \mathbf{x}_{1:l}^m, \nu^m, a, b) \propto p(\mathbf{x}_{1:l}^m \mid z^m) p(\nu^m \mid z, b) p(z^m \mid a) \tag{78}$$

and within the setup of Example 2, is again given by a Gamma distribution, i.e.,
$\mathcal{G}\left(\alpha + r_l^m + \rho, \beta + \int_0^{t_l} g_1(x^m(\tau))d\tau + \phi\nu^m\right)$. It follows that the marginal posterior distributions over the extrinsic factors Z^m can again be approximated by sums of Gamma distributions.

Implementation Aspects

Although the algorithm structure is mathematically characterized by the foregoing derivations, several implementation variants are possible. It turns out that certain implementation details may have significant impact on the achieved performance. For instance, resampling of the static parameters can be carried out before or after sampling the dynamic states. Both strategies are mathematically correct, however, the former generally achieves better results as every particle that is drawn from the distribution at t_{l-1} receives a newly sampled value. In contrast, when sticking to the latter strategy, some of the resampled parameters are immediately lost as the corresponding particles are never drawn again at the subsequent time step. Furthermore - unlike classical sequential importance sampling methods - the proposed algorithm requires a short burn-in period at each time iteration. In all simulation studies, we discarded around 10 percent of the particles.

Finally, we mention that one is free to chose the order of the recursive updates, meaning that single measurements can be incorporated first over time and then over cells or vice versa. If one is mainly interested in parameter estimation, we recommend to use the latter strategy, whereas the

sequence of processed cells at a particular time-step should be chosen randomly. This is likely to produce diversified summary statistics and in turn smooth posterior distributions over parameters. In contrast, processing entire cell trajectories one after each other appears to be beneficial if one aims to perform a state reconstruction. In the same context, we would like to discuss an interesting modification of the algorithm. In particular, it is based on the idea to resample and update individual cells simultaneously, i.e., without updating the posterior between consecutive cells. While a theoretical analysis of that algorithm shall be performed in the future, it has proven to perform well for both state reconstruction and parameter inference while getting along with comparably few particles per time instance. All of the three algorithm variants will be supported by the public DPP toolbox (i.e., *cells-first, time-points-first, simultaneous*).

An exemplary implementation of the marginal SMCMC algorithm (i.e., the *cells-first* variant) is summarized in Algorithm 1.

Algorithm 1 (Marginal SMCMC) *We assume that we have given the most recent posterior distribution as a set of particles. Then, the updated posterior distribution incorporating the next measurement of cell* m *is obtained by:*

1.) ***For each*** *particle* $p = 1, \ldots, P$*:*

   1.) *Select the* p*-th particle from the particle distribution at time* t_{l-1}.

   2.) *Resample the measurement parameters* $\tilde{\omega}$ *using (64) or (65).*

   3.) *Resample the extrinsic statistics and morphological shape parameters* $\tilde{\xi} = \{\tilde{\alpha}, \tilde{\beta}, \tilde{\rho}, \tilde{\phi}\}$ *using a M-H step with proposal density* $q(\tilde{\xi}) = \prod_{i=1}^{4} \mathcal{LN}\left(\ln \xi_i, \sigma_\xi^2\right)$ *and acceptance probability (72).*

   4.) *Propose a sub-trajectory* $\widehat{\mathbf{x}}_{l-1:l}^{m} \sim p(\mathbf{x}_{l-1:l}^{m} \mid x_{l-1}^{m}, \tilde{a}, \tilde{b}, \nu^{m}, T_{l-1}^{m})$ *by simulating the marginal dynamics on* $[l-1, l]$ *using the resampled parameters* $\tilde{a}$ *and* $\tilde{b}$.

   5.) *Merge the sub-trajectories to obtain a full sample path* $\widehat{\mathbf{x}}_{1:l}^{m} = \{\mathbf{x}_{1:l-1}^{m}, \widehat{\mathbf{x}}_{(l-1,l]}^{m}\}$.

   6.) ***If*** $p = 1$*: Accept particle with probability* 1.
   Else*: Accept particle with probability*

$$\gamma_x^m = \min\left\{1, \frac{p(y_l^m \mid \widehat{x}_l^m, \tilde{\omega})}{p(y_l^m \mid \tilde{x}_l^m, \omega)}\right\},$$

   where ω *denotes the measurement noise parameter of the previous particle.*

   7.) *Update* p*-th particle of the posterior distribution at time* t_l *using the newly sampled quantities.*

Remark: Note that the case where no morphological features are used for the inference can be understood as a special instance (or simplification) of the described algorithm. Hence, we do not provide additional equations for this scenario - in particular as they are straight-forward to obtain from the provided derivations.

Bayes Factor Computation

We perform model selection by calculating the *Bayes factor* for two competing models $\mathcal{M}_1$ and $\mathcal{M}_2$ with equal prior probability $P(\mathcal{M} = m_1) = P(\mathcal{M} = m_2) = 0.5$. We consider pooled time-course measurements from a heterogeneous population but for simplicity exclude the morphological features from the analysis. Note that in this case, only the parameters A and Ω are required to specify the model. The Bayes factor is then given by

$$\mathcal{K}_{1,2} = \frac{p(y^1_{1:l}, \ldots, y^M_{1:l} \mid m_1)}{p(y^1_{1:l}, \ldots, y^M_{1:l} \mid m_2)}. \tag{79}$$

Within the SMCMC framework, the marginal likelihood (i.e., the *model-evidence*) computations in (79) turn out to be straight forward as they can be carried out recursively as

$$\begin{aligned} &p(y^1_{1:l}, \ldots, y^M_{1:l} \mid m_k) \\ &\qquad = p(y^1_1, \ldots, y^M_1 \mid m_k) \prod_{i=2}^{N} p(y^1_i, \ldots, y^M_i \mid y^1_{1:i-1}, \ldots, y^M_{1:i-1}, m_k). \end{aligned} \tag{80}$$

The individual terms in (80), i.e., the predictive densities, are given by

$$p(y^1_i, \ldots, y^M_i \mid y^1_{1:i-1}, \ldots, y^M_{1:i-1}, m_k) = \mathbb{E}\left[\prod_{m=1}^{M} p(y^m_i \mid x^m_i, \omega)\right] \tag{81}$$

where the expectation is take with respect to the density
$p(\mathbf{x}^1_{1:i}, \ldots, \mathbf{x}^1_{1:i}, a, \omega \mid y^1_{1:i-1}, \ldots, y^M_{1:i-1})$, which is simply obtained by drawing a particle
$\{\mathbf{x}^1_{1:i-1}, \ldots, \mathbf{x}^1_{1:i-1}, a, \omega\}^{(p)}$ from the previous time step and extending the dynamic states until t_i using the parameters $a^{(p)}$ and $\omega^{(p)}$ from that particle.

Bayes factors are often expressed in units of information such as in Deciban (dB), i.e.,

$$\mathcal{K}_{1,2} = \log_{10}\left(\frac{p(y^1_{1:l}, \ldots, y^M_{1:l} \mid m_1)}{p(y^1_{1:l}, \ldots, y^M_{1:l} \mid m_2)}\right) \mathrm{dB}. \tag{82}$$

Model-based Noise Decomposition

Let Y_l be a one-dimensional noisy readout of the X_l at time t_l and Z a constant extrinsic factor. For the sake of clarity, all other model quantities

are dropped in the notation of the following derivation. The total variance of Y_l can be decomposed as

$$\begin{aligned}
\operatorname{Var}[Y_l] &= \mathbb{E}\left[Y_l^2\right] - \mathbb{E}[Y_l]^2 \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[Y_l^2 \mid X_l\right] \mid Z\right]\right] - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2 \\
&= \mathbb{E}\left[\mathbb{E}\left[\operatorname{Var}[Y_l \mid X_l] + \mathbb{E}[Y_l \mid X_l]^2 \mid Z\right]\right] - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2 \\
&= \mathbb{E}[\mathbb{E}[\operatorname{Var}[Y_l \mid X_l] \mid Z]] + \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[Y_l \mid X_l]^2 \mid Z\right]\right] \\
&\qquad - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2 \\
&= \mathbb{E}[\mathbb{E}[\operatorname{Var}[Y_l \mid X_l] \mid Z]] \\
&\qquad + \mathbb{E}\left[\operatorname{Var}[\mathbb{E}[Y_l \mid X_l] \mid Z] + \mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]^2\right] \\
&\qquad - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2 \\
&= \mathbb{E}[\mathbb{E}[\operatorname{Var}[Y_l \mid X_l] \mid Z]] + \mathbb{E}[\operatorname{Var}[\mathbb{E}[Y_l \mid X_l] \mid Z]] \\
&\qquad + \mathbb{E}\left[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]^2\right] - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2 \\
&= \mathbb{E}[\mathbb{E}[\operatorname{Var}[Y_l \mid X_l] \mid Z]] + \mathbb{E}[\operatorname{Var}[\mathbb{E}[Y_l \mid X_l] \mid Z]] \\
&\qquad + \operatorname{Var}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]] \\
&\qquad + \underbrace{\mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2 - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]^2}_{=0} \\
&= \underbrace{\mathbb{E}[\mathbb{E}[\operatorname{Var}[Y_l \mid X_l] \mid Z]]}_{\text{technical}} + \underbrace{\mathbb{E}[\operatorname{Var}[\mathbb{E}[Y_l \mid X_l] \mid Z]]}_{\text{intrinsic}} \\
&\qquad + \underbrace{\operatorname{Var}[\mathbb{E}[\mathbb{E}[Y_l \mid X_l] \mid Z]]}_{\text{extrinsic}}.
\end{aligned} \tag{83}$$

Note that since the SCV is defined as the ratio between the variance and squared mean, the same decomposition applies, whereas each of the three contributions is divided by a time-dependent factor, i.e., the squared expectation of the measurement Y_l. All three terms were computed via simulation of the calibrated model, by setting the model parameters to their mean posterior estimate.

SUPPLEMENTARY NOTE 2:

INFERENCE RESULTS USING SIMULATED DATA

The gene expression model considered in this case study is depicted in Figure 25, where $u(t)$ denotes a time-dependent gene-activation rate (e.g., driven by a temporary TF-translocation) and $c_2 - c_6$ denote stochastic rate constants. All reactions are assumed to follow mass-action kinetics.

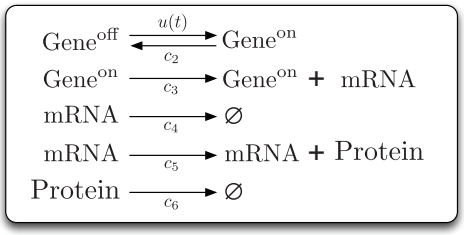

Figure 25: Kinetic model for two-stage gene expression.

The input rate $u(t)$ is chosen to switch between two values (i.e., $u_0 = 0s^{-1}$ and $u_1 = 3.00e-05s^{-1}$), such as indicated in Figure 14 C in the main text. For simplicity we assume $u(t)$ and $c_6 = 4.00e-04s^{-1}$ to be known. Kinetic parameters c_2, c_3 and c_4 are to be inferred from the measurements, whereas we assume independent and Gamma-distributed prior knowledge. Extrinsic heterogeneity is modeled in the translation efficiency (i.e., $z \equiv c_5$). More specifically, we assume a heterogeneity of the form $Z \mid (A = a) \sim \mathcal{G}(\alpha, \beta)$. Each cell being measured comes with an additional morphological feature (e.g., such as the cell size), which is assumed to statistically depend on the translation efficiency such that $V \mid (Z = z, B = b) \sim \mathcal{G}(\rho, \phi z)$. The parameters $b = \{\rho, \phi\}$ describing this dependency together with $a = \{\alpha, \beta\}$ are as well estimated from the data. For each of those quantities, we assume independent log-normal prior distributions. Similarly, we estimate the measurement noise parameter (i.e., the scaling parameter of a log-normal distribution) with a Gamma-shaped prior uncertainty. Statistics of the prior and resulting posterior distributions are given in Table 5. A density plot of the posterior distribution over the morphological parameters ρ and ϕ is depicted in Figure 26.

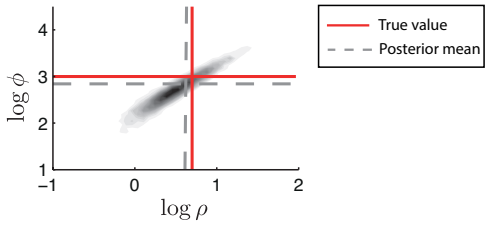

Figure 26: Posterior density plot over the morphological parameters ρ and ϕ.

			Prior			Posterior		
Name	Reference	Unit	q_5	q_{95}	Mean	q_5	q_{95}	Mean
c_2	$5.00e-03$	s^{-1}	$3.40e-04$	$2.03e-02$	$6.67e-03$	$3.79e-03$	$6.85e-03$	$5.23e-03$
c_3	$1.00e+00$	s^{-1}	$5.13e-02$	$2.98e+00$	$1.00e+00$	$8.43e-01$	$1.18e+00$	$1.01e+00$
c_4	$2.00e-02$	s^{-1}	$1.78e-03$	$9.72e-02$	$3.33e-02$	$1.70e-02$	$2.67e-02$	$2.17e-02$
α	$3.00e+00$	$-$	$2.62e-01$	$7.16e+00$	$2.25e+00$	$5.26e-01$	$5.10e+00$	$2.22e+00$
β	$1.00e+02$	$-$	$3.20e+01$	$8.54e+02$	$2.71e+02$	$1.47e+01$	$1.28e+02$	$5.39e+01$
ρ	$5.00e+00$	$-$	$7.04e-01$	$7.81e+01$	$1.96e+01$	$9.14e-01$	$1.09e+01$	$4.19e+00$
ϕ	$1.00e+03$	$-$	$7.74e+01$	$7.32e+03$	$1.98e+03$	$1.22e+02$	$1.91e+03$	$6.93e+02$
ω	$1.50e-01$	$-$	$6.50e-02$	$2.36e-01$	$1.26e-01$	$1.24e-01$	$1.57e-01$	$1.39e-01$

Table 5: Inferred model parameters and credible intervals for the synthetic case study.

SUPPLEMENTARY NOTE 3:

STRAINS AND PLASMIDS

Plasmids and Yeast strains are listed in Tables 6 and 7.

pSP45 and pSP212 are based on pRS vectors [207]. pSP45 was constructed by inserting a Ubi-Y destabilizing sequence between SpeI and HindIII and a single Venus fluorescent protein cassette between HindIII and XhoI. pSP212 was cloned by inserting an mCherry protein between XbaI and HindIII in a pRS315 pTEF plasmid. The GEV sequence was inserted between the promoter and the fluorescent protein by gap repair.

For yMU16, plasmids pSP45 and pSP212 were integrated into ySP37, where the nuclear protein Hta2 is tagged with CFP by adapter mediated genome alteration [186] in a Wild Type (WT) yeast *Saccharomyces cerevisiae* (W303) (ySP2). Colonies of strong and homogeneous expressing transformants were selected by microscopy after induction with β-estradiol after each transformation step.

The GEV-mCherry was integrated in the URA3 locus by homologous recombination between a pRS306 plasmid cut in the Multiple Cloning Site (MCS) site and a PCR amplifying the MCS region of pSP212.

yMU19-28 were constructed by tagging the respective protein in the ySP2 WT strain with a Venus protein (pKT90) [204]. Strong and homogeneous expressing transformants were selected by microscopy.

Plasmid	Description	Source/Ref
pSP45	pRS305 pGAL1-Ubi-destabY-Venus	This study
pSP94	pFa6A mCherry-Ura	This study
pSP212	pRS415 pTEF-GEV-mCherry	This study

Table 6: List of Plasmids

Strain	Genotype	Source/Ref
ySP2	MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15	QUASI consortium
ySP37	MATa hta2::hta2-CFP	This study
yMU16	MATa hta2::hta2-CFP leu2::LEU2-pGAL1-Ubi-destabY-Venus ura3::URA3-pTEF-GEV-mCherry	This study
yMU19	hog1::hog1-Venus-SpHIS5	This study
yMU21	car1::car1-Venus-SpHIS5	This study
yMU22	gpx2::gpx2-Venus-SpHIS5	This study
yMU23	hsp104::hsp104-Venus-SpHIS5	This study
yMU24	sse2::sse2-Venus-SpHIS5	This study
yMU25	tda1::tda1-Venus-SpHIS5	This study
yMU26	tma108::tma108-Venus-SpHIS5	This study
yMU28	ygr117c::ygr117c-Venus-SpHIS5	This study
ySP247	MATa leu2::LEU2-pGAL1-quadruple-Venus his3::HIS3-pGAL1-quadruple-CFP trp1::Trp1-pADH-GEV	This study

Table 7: List of Yeast Strains

SUPPLEMENTARY NOTE 4:

QUANTIFICATION OF PROTEIN LEVELS

A mapping between fluorescence intensities, obtained by microscopy experiments, and the absolute copy numbers of fluorescent proteins was established. From proteins in the Yeast GFP Fusion Localization Database [104] that showed a homogenous distribution in the cytoplasm, we selected several ones with numbers of molecules per cell [75] covering approximately the range from 1000 to 40000. Selected proteins and the numbers of molecules per cell are listed in Table 8

Standard Name	Systematic Name	Molecules per Cell
	Ygr117cp	1280
Gpx2p	Ybr244wp	2010
Tma108p	Yil137cp	5110
Sse2p	Ybr169cp	6300
Hog1p	Ylr113wp	6780
Tda1p	Ymr291wp	10200
Hsp104p	Yll026wp	32800
Car1p	Ypl111wp	42800

Table 8: List of Reference Proteins

Strains yMU19-28, each with one of the reference proteins tagged with a single Venus fluorescent reporter, were constructed as described in Supplementary Note 3. A WT strain ySP2 (to account for the autofluorescence) and the reference strains yMU19-28 were then grown and prepared before 200 µl of the diluted and briefly sonicated cell solution were loaded into 96-well plates for obtaining fluorescence microscopy images (see Supplementary Note 3).

A mapping between the total fluorescence intensities, averaged over several cells, and absolute protein copy numbers was obtained by a linear regression model of the form

$$I = kn + d_{AF} + \xi, \tag{84}$$

with I as the total intensity within a cell, k as a scaling constant, n as the absolute number of proteins, d_{AF} as a constant bias arising from autofluorescence and ξ as a zero-mean, normally distributed error term. For a graphical illustration and calibrated single cell traces see Figure 27 and Figures 20 and 21.

An aggregation of sequential events downstream of the transcription initiation process (e.g., post-transcriptional / translational modifications, reporter maturation, etc.) may induce a static delay in the imaging-based Y-Venus readout. This effect is also visible in Figure 27, where the protein

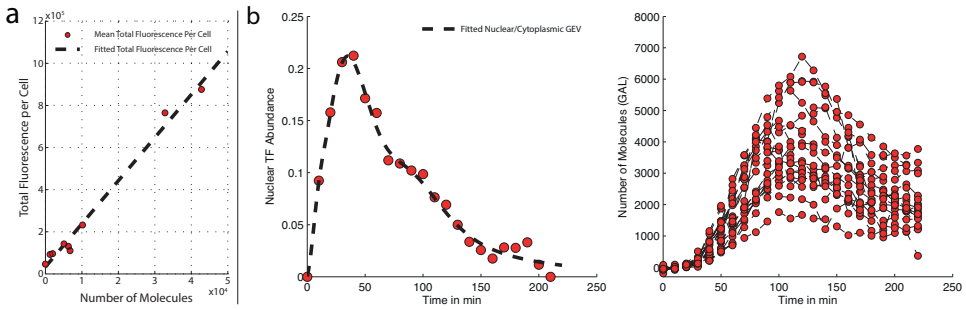

Figure 27: Quantification of protein levels. (a) Linear regression (dashed line) between the total fluorescence per cell and the number of molecules. The red circles show the total YFP fluorescence per cell averaged over several cells for measured strains (ySP2, yMU19-28). (b) Calibrated single cell traces of Y-Venus expression over time for a 50 nM pulse of β-estradiol.

readout appears to fluctuate around zero for the first few time points. In order to correct the quantified protein traces by that delay, we apply a one-sided Kolmogorov-Smirnov test between the protein samples at time $t = 0\,\text{min}$ and the subsequent acquisition points. Hence, the test decides whether the protein abundance over all cells at time points $t_1, t_2, \ldots$ is statistically higher than that of $t_0 = 0\,\text{min}$. The resulting p-values for the increasing time-delays and the desired confidence level (i.e., $p = 0.05$) are depicted in Figure 28. The p-value significantly drops for a time delay of $\Delta t = 30\,\text{min}$ and hence, we removed the first three time-points of the measured abundance traces.

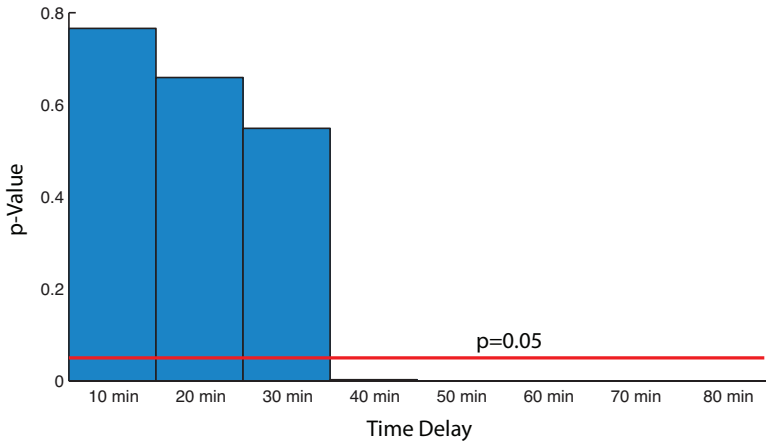

Figure 28: Time-delay estimation in Y-Venus abundance measurements.

We remark that several alternative strategies to account for the time-delay could be followed. For instance, one can include maturation reactions for the fluorescent proteins or include the time-delay as an additional unknown parameter into the SMCMC scheme or estimate the initial conditions at the first non-removed time point - to name a few. Furthermore, one could explicitly model additional multi-step reactions that are likely

to be present at different stages of gene expression. However, the reader should keep in mind that including more unknown parameters into the inference will yield an increased overall estimation uncertainty.

SUPPLEMENTARY NOTE 5:

INFERENCE RESULTS FOR THE β-ESTRADIOL EXPRESSION SYSTEM

We performed a Bayesian model selection as described in *Supplementary Note 1* in order to find out which of the three hypothesized kinetic models from Figure 19 is best supported by the experimental data. Using the same procedure, we also decided between a normally and log-normally distributed measurement noise. Note that upon execution of the DPP-based inference, no extra computational effort is needed, as the required statistics are retrieved within the SMCMC algorithm. In order to obtain a fair comparison, we chose parameter prior information being roughly consistent across all the models under study. High evidence was found for a log-normally distributed measurement error (i.e., above 100 dB compared to normally distributed error in conjunction with model **a**). Note that the obtained model rankings are subject to sampling variance. Hence, in case of the three kinetic models, we computed the model-evidences five times in order to check the robustness of the obtained results. We know from *Supplementary Note 1* that the evidence is computed as a product of predictive densities, each corresponding to a particular time-point. Hence, in order to approximately assess the variability over different runs, we randomly combined the individual terms over the five repeats and computed histograms over the resulting model-evidences. The results (shown in Figure 29) demonstrate that the two-state model **a** ranked best. Although there is a significant overlap with the histogram of the three-state model **c**, the results indicate no need for using the more complicated model **c**. Moreover, the figure indicates little evidence for model **b** compared to **a** and **c**. We then computed the average model-evidence and the respective Bayes factors, i.e., around 6.8 dB when comparing **a** to **c** and 51.8 dB when comparing **a** to **b**. For compactness, prior and posterior statistics are only given for the winning model **a** (see Table 9).

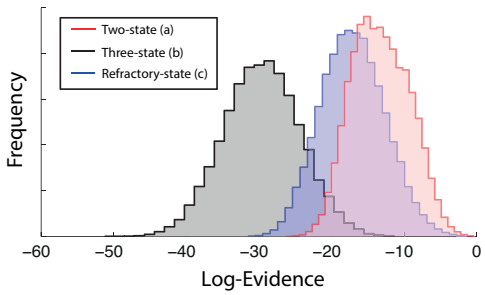

Figure 29: Robustness evaluation of Bayesian model selection with respect to models **a**, **b** and **c**. All histograms were shifted by the same constant for a better visualization.

For c_1, c_2 and c_3 we used weakly informative exponential priors, i.e., $p(\cdot) = \mathcal{G}(1, 10)$, with their quantiles shown in Table 9. In case of the

mRNA and protein degradation rates c_4 and c_6, the priors were chosen such that roughly 95 percent of the probability mass were within the ranges $[3, 40]$ and $[15, 2000]$ minutes expected half life (i.e., $C_4 \sim \mathcal{G}(3, 2000)$ and $C_6 \sim \mathcal{G}(1, 5000)$). Furthermore, for the hyperparameters α and β, a two-dimensional log-normal prior distribution $\mathcal{LN}(\mu_A, \Sigma_A)$ with

$$\mu_A = \begin{pmatrix} 2.71 \\ 5.70 \end{pmatrix} \text{ and } \Sigma_A = \begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.20 \end{pmatrix}$$

was used such as to be consistent with the results found in [166]. In order to avoid a biased inference, we again picked a weakly informative log-normal prior distribution over the morphological parameters ρ and ϕ, i.e., $\mathcal{LN}(\mu_B, \Sigma_B)$ with

$$\mu_B = \begin{pmatrix} 2.30 \\ 10.82 \end{pmatrix} \text{ and } \Sigma_B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Furthermore, Figure 30 shows the posterior densities for each pair of kinetic rate constants.

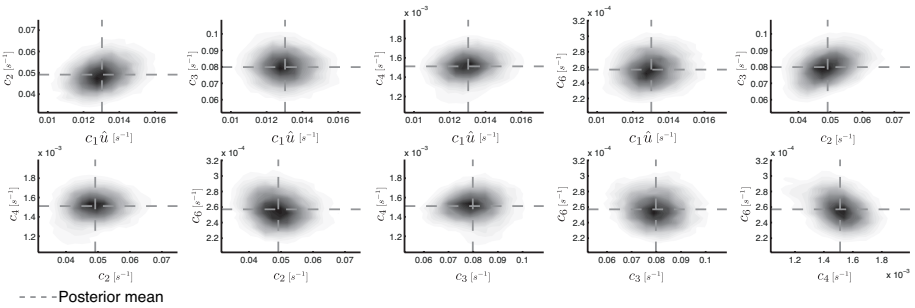

Figure 30: Inferred posterior densities between individual rate constants. As in the main text, $\hat{u}$ denotes the temporal average of the nuclear GEV intensity; the mean values of the posterior distributions are shown by dashed lines.

In order to asses the sensitivity of the inferred gene-switching rates with respect to the prior distribution, we re-ran the inference procedure using a weaker prior distribution. In particular we used for both c_1 and c_2 a Gamma distribution $\mathcal{G}(0.5, 5)$, showing heavier tails as well as more probability mass for small values of c_1 and c_2 (i.e., slow switching). Figure 31 shows a density plot of the prior and the obtained posterior distribution. The resulting mean posterior estimates are practically equivalent to the results from Table 9. Although the prior density takes comparably small values for high switching rates, the posterior was pushed to that parameter region.

MCMC DIAGNOSTICS. The algorithm is based on a sequential execution of individual MCMC samplers (corresponding to time points and cells).

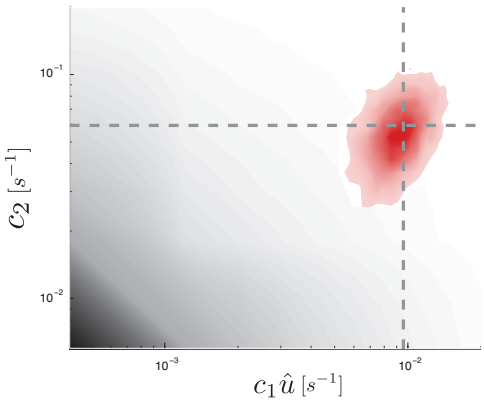

Figure 31: Inferred posterior density (red) over gene-on and gene-off rate using weak priors (gray-shaded), favoring slow switching; posterior mean estimate shown by dashes lines.

One strategy to evaluate the sampling quality of MCMC schemes is to determine their *effective number of samples*. This number estimates how many independent samples have been produced by the Markov chain and hence helps to find a reasonable number of algorithm iterations. The (relative) effective number of samples is given by

$$r_{eff} = \frac{1}{1 + 2\sum_{i=1}^{\infty} \rho_k} \tag{85}$$

with ρ_k as the k-lag autocorrelation function of the MCMC output. Figure 32 shows the average effective number of samples of the sampled Y-Venus abundance over the measurement times, whereas the average is taken over all cells. The figure indicates that sampling is particularly difficult at the first time point, where only little is known about the underlying process and its parameters. Hence, we decided to use twice as many iterations during the first iteration to obtain at least around 1000 independent samples.

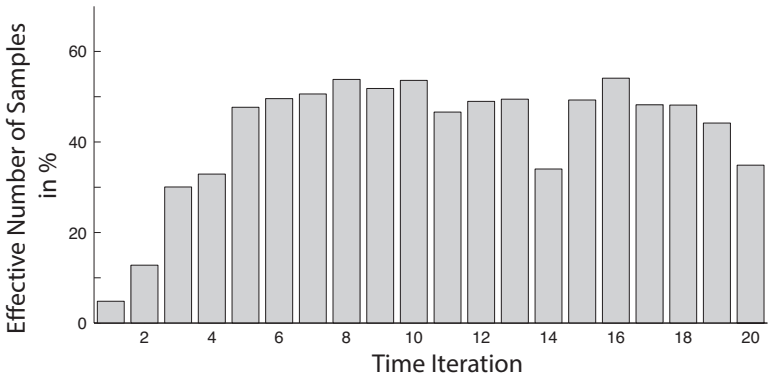

Figure 32: Effective number of samples for the used inference scheme. Each bar denotes the average taken over all cells.

		Prior			Posterior		
Name	Unit	q_5	q_{95}	Mean	q_5	q_{95}	Mean
c_1	s^{-1}	$5.35e-03$	$3.03e-01$	$1.00e-01$	$1.08e-01$	$1.41e-01$	$1.24e-01$
c_2	s^{-1}	$5.40e-03$	$3.03e-01$	$1.00e-01$	$3.98e-02$	$5.93e-02$	$4.92e-02$
c_3	s^{-1}	$5.00e-03$	$2.96e-01$	$1.00e-01$	$6.72e-02$	$9.27e-02$	$7.99e-02$
c_4	s^{-1}	$4.14e-04$	$3.16e-03$	$1.50e-03$	$1.30e-03$	$1.71e-03$	$1.51e-03$
c_6	s^{-1}	$1.13e-05$	$6.07e-04$	$2.00e-04$	$2.30e-04$	$2.87e-04$	$2.57e-04$
α	–	$8.78e+00$	$3.83e+01$	$2.04e+01$	$3.87e+00$	$3.83e+01$	$1.41e+01$
β	–	$1.60e+02$	$1.56e+03$	$6.29e+02$	$3.88e+01$	$4.01e+02$	$1.45e+02$
ρ	–	$7.38e+00$	$7.84e+02$	$2.05e+02$	$7.87e-01$	$1.42e+01$	$4.72e+00$
ϕ	–	$3.52e+04$	$3.98e+06$	$1.02e+06$	$6.69e+04$	$3.32e+06$	$1.02e+06$
ω	–	$8.94e-02$	$2.44e-01$	$1.48e-01$	$1.13e-01$	$1.38e-01$	$1.25e-01$

Table 9: Inferred model parameters and credible intervals for the β-estradiol induced gene expression system.

SUPPLEMENTARY NOTE 6:

COMPARISON TO PREVIOUS APPROACHES

Stochasticity in gene expression has attracted significant attention during recent years. Accordingly, many analytical and experimental techniques have been developed to study the role of noise in promoter-, transcriptional- and translational dynamics. In this section we aim to give a brief overview about related approaches and outline how they differ from our method. Most analytical approaches that have been successfully applied to experimental data rely on derivations of mRNA and protein distributions under certain mathematical assumptions. For instance, a bursting model of protein synthesis was proposed in [69], where the authors have shown that at stationarity, proteins are Gamma-distributed if one assumes a kinetic model of the form

$$\begin{aligned} \emptyset &\underset{c_2}{\overset{c_1}{\rightleftharpoons}} \text{mRNA} \\ \text{mRNA} &\xrightarrow{c_3} \text{mRNA} + P \\ P &\xrightarrow{c_4} \emptyset \end{aligned} \tag{86}$$

in conjunction with an approximately exponentially distributed protein burst-size. The clear advantage of that approach is its analytical simplicity. In fact, the derived distribution can be easily fitted to experimentally obtained protein distributions. The distribution parameters a and b of that Gamma distribution can be used to characterize the expression dynamics: $a = c_1/c_4$ gives the average *burst frequency* and $b^{-1} = c_3/c_2$ is the average *burst size*[3]. This approach was for instance applied in [69], where the authors have demonstrated that the Gamma distribution represents a suitable steady-state approximation for many proteins in *E. coli*.

A conceptually related method was proposed in [180], where steady-state distributions were derived for mRNA abundance assuming a kinetic model of the form

$$\begin{aligned} I &\underset{c_2}{\overset{c_1}{\rightleftharpoons}} A \\ A &\xrightarrow{c_3} A + \text{mRNA} \\ \text{mRNA} &\xrightarrow{c_4} \emptyset, \end{aligned} \tag{87}$$

where I and A refer to the molecular states where the promoter is inactive or active. Similar to [69], average bursting kinetics at the transcription level can be related to the parameters of the kinetic model: c_3/c_2 and c_1/c_4 give the average burst-size and frequency, respectively. The transcriptional dynamics of different genes can then be characterized according to such measures. For instance, genes with an average burst-size above or below one are said to be *bursty* or *non-bursty*, respectively.

3 Note that the parameter b of the Gamma-distribution is inversely defined in [69]

While undoubtedly useful, both described approaches are characterized by the following limitations: First, the derived formulas do not apply for non-stationary processes, which is the case for any transiently induced expression system (such as considered in this work). Second, in the context of parameter inference, both approaches cannot resolve individual rate constants but only ratios thereof (e.g., such as the average burst-size). Since any temporal information of the underlying process is lost, such models exhibit a so-called *structural non-identifiability*. This means that even if one assumes complete and noise-free measurements of the distribution, the individual parameters cannot be resolved[4]. While the authors in [180] consider the particular values of that parameters to be of minor importance for their study, we argue that assessing those values can be of vital importance when characterizing expression kinetics. More specifically, they can resolve the absolute time-scale on which transcription and translation take place.

Recent approaches that can deal with transient distribution data include [256, 165]. As they were designed for population snapshot data (e.g., such as mRNA FISH or flow cytometry) they do not include temporal information on a *single-cell level*, such as provided by time-lapse microscopy data. Using such an approach on the data we provide, one neglects a significant portion of information, which in turn leads to less accuracy (or possibly even non-identifiabilities) in the resulting parameter inference. As a consequence, biophysical quantities such as gene-switching times or mRNA bursting statistics are predicted less accurately. Dealing with stochastic models, there exists a lower bound on the prediction uncertainty that just corresponds to the process uncertainty itself (i.e. present even in the case of complete knowledge of the parameters). Any positive deviation from this bound stems from further parameter uncertainty represented by the posterior distribution. Assuming finite data records, there exist again fundamental lower bounds for this uncertainty (e.g., the minimum mean squared error). While some inference methods may exploit all features of the data, some others may not. Accordingly, the former will achieve this minimal posterior uncertainty bound while the latter will not. Hence, one anticipates a higher variance in the predictive distribution for methods that do not exploit the temporal correlation structure in the data. We performed a simple case study using the transcriptional model from eq. (87) to confirm this expectation (see Fig. 33).

The present approach was designed to fully exploit the information revealed by pooled time-lapse microscopy measurements – without relying on mathematical approximations of the biochemical process. On the one hand this implies several technical difficulties, since the full temporal CTMC dynamics need to be reconstructed for every considered cell. On the other hand – however – the approach can exploit the information across time-points and cells within an experiment. As a consequence, it can re-

4 In contrast, *practical non-identifiabilities* can be tackled by including more data such as demonstrated in section *Supplementary Note 8*

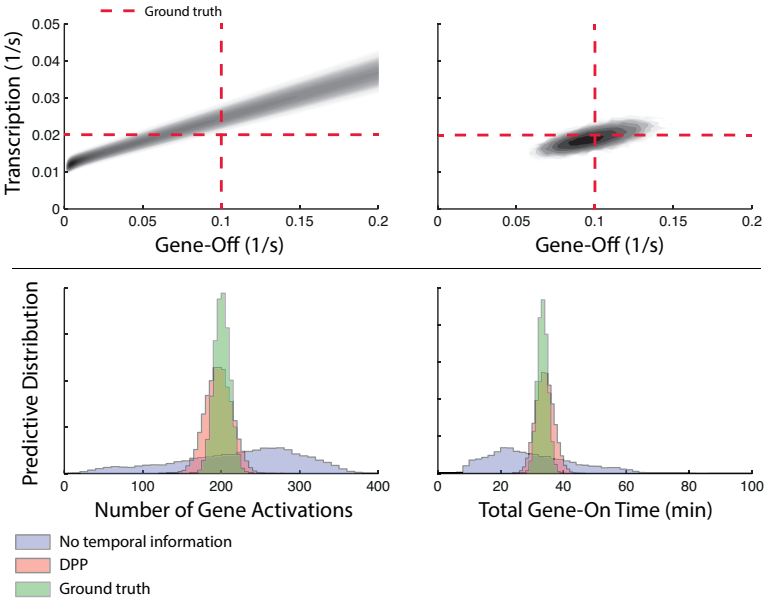

Figure 33: Predictions obtained using an inference scheme that neglects temporal information within single-cells (e.g., [165]) and DPP. Measurements of 20 cells were simulated at 20 equally spaced time points between 0 and 100 min using a log-normal measurement model with $\sigma = 0.15$. Inference was performed with respect to gene-off- and transcription rates. The upper panel shows the respective posterior distributions (left: population snapshot, right: DPP). The lower panel shows the predictive distributions for two quantities characterizing the underlying process: (left) the total number the gene is activated within 100 min. (right) the total time the gene is in its active state within 100 min. The predictions are compared to the distributions obtained using the true kinetic parameters (ground truth).

solve non-identifiabilities and allows a joint and approximation-free quantification of kinetic parameters and the population's heterogeneity.

Nevertheless, in order to link our computational predictions to existing work, we compute the aforementioned measures using our inferred parameter values. For instance, we obtain for the average burst-size $c_3/c_2 = 1.6$ indicating that mRNA is transcribed in (small) bursts. However, The authors from [260] point out that the burst-size alone does not paint a full picture of the different modes of transcription observed experimentally. Instead, they suggest another two-dimensional characterization, i.e., the rate of transcription (i.e., c_3) together with a value defined by $\text{fraction}^{-1} = (c_1 + c_2)/c_1$. The latter gives the inverse of the relative time when the gene is active. The authors have shown that experimentally evidenced expression models will then be scattered around a line with slope $c_3 c_1/(c_1 + c_2)$ defining the *effective* rate of transcription – taking into account also the time where the gene is inactive. Using our parameter estimates, these values are given by $c_3 = 4.8\,\text{min}^{-1}$ and $\text{fraction}^{-1} = 4.8$ and hence, agree very well with figures 7 and 8 from [260].

Furthermore, our state-reconstruction yielded maximum mRNA levels in the order of $20-50$ transcripts per cell, being consistent with the results from [165], where mRNA distributions where directly measured by mRNA FISH in yeast.

SUPPLEMENTARY NOTE 7:

COMPARISON BETWEEN HOMOGENEOUS AND HETEROGENEOUS MODELS

Most state-of-the-art approaches for parameter inference in biochemical networks do not account for extrinsic variability – that means – they rely on the assumption that the recorded measurements stem from a homogeneous cell population. If the considered biochemical system is characterized by significant heterogeneity, such models and their respective inference will yield biased results – either in the resulting parameter estimates or subsequent predictions. In order to demonstrate this issue, we refitted the model under the assumption of homogeneity using the same prior distributions over the kinetic and measurement parameters. Subsequently, that model was used to predict the three pulse experiments. Fig. 34 shows the predicted and experimental means and standard deviations of the Y-Venus abundance for the different concentrations. For completeness, also the predictions from the original (i.e., heterogeneous) model are shown. We find that in this case, a model that neglects extrinsic variability is not able to explain the large variability present in the data. While the mean dynamics are captured well for all concentrations, the standard deviation is consistently underestimated, highlighting the importance of inference schemes that can account for extrinsic variability.

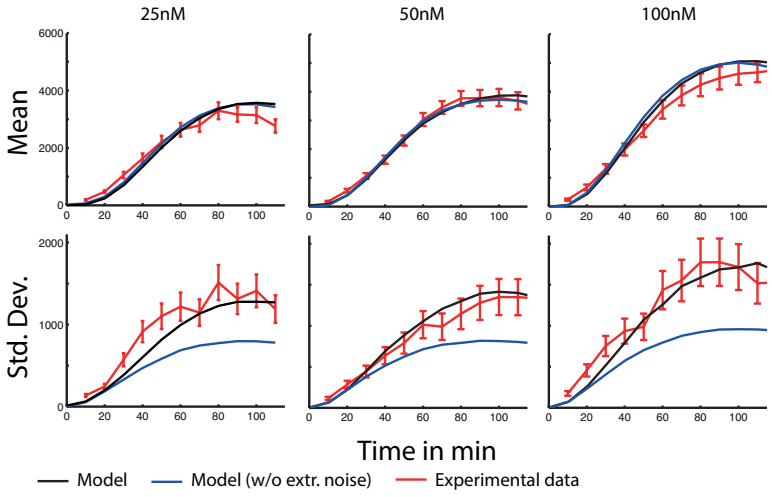

Figure 34: Predictions obtained using an inference scheme that neglects extrinsic variability. The results are compared to the experimental data as well as to the model predictions obtained via DPP. Wiskers indicate standard errors of the experimentally obtained quantities.

SUPPLEMENTARY NOTE 8:

IMPROVED IDENTIFIABILITY VIA POOLED RECORDINGS

Joint inference of multiple rate constants from single trajectories often yields practical non-identifiabilities. However, the ill-posedness of such problems can be drastically reduced by pooling recordings over multiple cells. This is demonstrated in the following simulation study where two parameters of the two-state gene expression model (i.e., c_2 and c_4) are jointly estimated using (i) one and (ii) ten single-cell trajectories. For both rate constants, we assumed prior distributions of the form $\mathcal{G}(1, 10)$. Density plots of the prior and the posteriors for case (i) and (ii) are depicted in Figure 35.

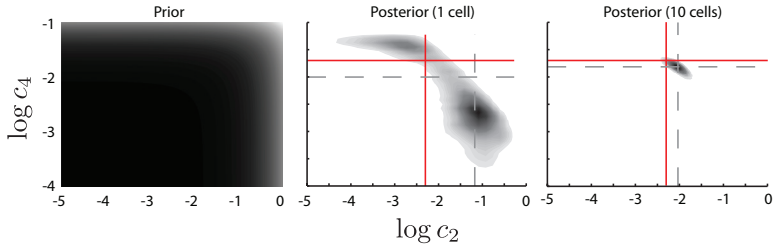

Figure 35: Improved identifiability via pooled single-cell trajectories; prior distribution (left), posterior distribution based on data from one cell (middle) and posterior based on ten cells (right); true parameter values (red lines) and posterior mean parameter estimates (dashed lines).

The left panel of Figure 35 demonstrates a weakly informative prior distribution. As a result, a single-cell is hardly enough to jointly identify both parameters, i.e., the posterior mass is distributed over roughly 4 orders of magnitude in both dimensions. When increasing the number of measured trajectories to ten, this non-identifiability is widely resolved such that both parameters can be inferred accurately.

SUPPLEMENTARY NOTE 9:

NUMERICAL COMPARISON BETWEEN DPP AND STANDARD RESAMPLING TECHNIQUES

We compared the proposed DPP-based inference scheme to a state-of-the-art resampling approach [219], where an invariant kernel is used to maintain diversity among particles. In particular, we computed the Mean Squared Error (MSE) between the mean posterior value of a single parameter from the corresponding true value. The simulation study was performed for a simple mass-action system, defined by the reactions shown in Figure 36, where $c_1 = 0.1, c_2 = 0.03$ and $c_4 = 0.001$ were assumed to be known, while $c_3 = 0.01$ was estimated using the two SMCMC approaches (with $P = 200$ particles) assuming a weakly informative prior distribution $C_3 \sim \mathcal{G}(0.01, 1)$.

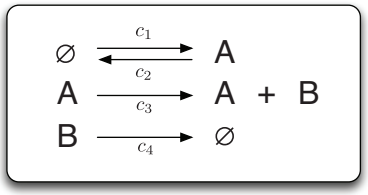

Figure 36: Simple mass-action model used for DPP performance evaluation.

For each run, we simulated a sequence of 15 measurements of species B at equally-spaced time points in the interval $[0\,\text{min}, 100\,\text{min}]$ with log-normally distributed acquisition noise (with known scaling parameter $\omega = 0.15$). We simulated $K = 1000$ runs for both approaches and computed the MSE μ_i and its standard deviation σ_i at the i-th time instance via bootstrapping. In particular, we randomly subsampled 10000 samples of size 800 and empirically determined the error statistics. The results are depicted in Figure 37.

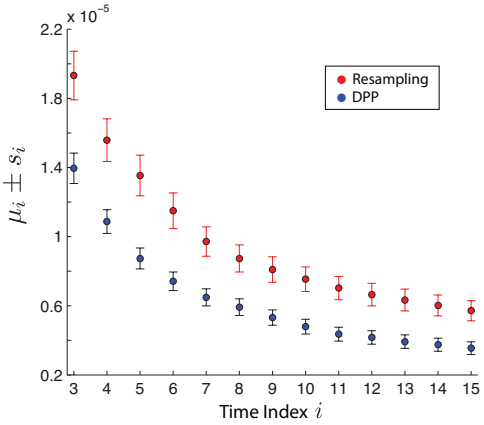

Figure 37: Performance comparison between DPP and standard resampling approach.

7

MANUSCRIPT 5

REFERENCE

M. Unger, C. Zechner, S. S. Lee, S. Pelet, M. Peter, and H. Koeppl, *Optimizing Single Cell Recordings for the Inference of Transcription Dynamics*, In Preparation

NOTE

The following manuscript summarizes the current state of the project, and is not yet ready for submission for publication.

AUTHOR CONTRIBUTIONS

MU performed experiments and wrote the manuscript. MU, SSL designed, developed and fabricated microfluidic devices. MU, CZ developed the mathematical model and performed simulations. MU, SP designed and constructed strains and plasmids. MU, CZ, SP, MP and HK designed the project.

OPTIMIZING SINGLE CELL RECORDINGS FOR THE INFERENCE OF TRANSCRIPTION DYNAMICS

Michael Unger[1,2], Christoph Zechner[1], Sung Sik Lee[2], Serge Pelet[3], Matthias Peter[2], and Heinz Koeppl[4]

ABSTRACT

Models of intracellular processes are often characterized by a large degree of uncertainty. Traditional experimental protocols tend to focus on intuitive interpretation of the results, or are limited by technical means, leading to limited excitation of the system under study, and consequently, to data with low information content for modeling tasks. In the following we present a novel microfluidic platform to generate complex concentration stimuli and record more informative single cell time-course data. We demonstrate how the selection of input stimuli critically influences the information content of a dataset. We apply an iterative Bayesian framework for the inference of kinetic parameters to a transcriptional network, controlled by the stress activated MAPK Hog1.

INTRODUCTION

Generating predictive computational models of intracellular processes is a task at the core of quantitative biology. To capture the stochastic nature and cellular heterogeneity many processes – such as gene expression mechanisms – exhibit[61], single cell measurements provide an informative data source for the statistical inference of network structures or kinetic model parameters [259, 258]. Although both experimental and computational tools made significant advances in recent years, their full potential often remains unexploited as individual methods are not carefully matched.

Applying optimal experimental design strategies [39] to biochemical experiments can help to close this gap and maximizing the information content [136] of a dataset for a specific modeling task. Rather than for direct interpretation, experimental variables, such as chemical perturbation sequences or acquisition time points, are then chosen to fulfill a specific objective function, such as optimally reducing inferred parameter uncertainties. In ref. [18], Bandara and colleagues showed how temporal concentration profiles of an inducer and inhibitor could be optimized to yield a significant reduction in parameter variance for a model of differential equations. Only recently, such studies emerged for models using stochas-

1 Automatic Control Laboratory, ETH Zurich, Zurich, Switzerland.
2 Institute of Biochemistry, ETH Zurich, Zurich, Switzerland.
3 Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland.
4 TU Darmstadt, Darmstadt, Germany. Correspondence should be addressed to H.K. (heinz.koeppl@bcs.tu-darmstadt.de)

tic chemical kinetics [120, 162, 256, 189]. Experiment design strategies have further been applied to model selection applications [33, 208].

Here we develop a model for stress induced transcription, controlled by the Mitogen-Activated Protein Kinase (MAPK) Hog1, and demonstrate how various perturbation sequences result in different reductions of parameter uncertainties. We further introduce a novel microfluidic platform to synthesize temporal perturbation profiles during time-lapse microscopy experiments.

RESULTS

Recursive Bayesian Experimental Design

We developed a suite of tools (Fig. 38) for iteratively optimizing single cell time-lapse assays for the inference of stochastic reaction kinetics, and thus for generating and refining predictive computational models. This includes a microfluidic platform that allows us to capture individual cells, and expose them to input sequences, i.e. temporal profiles of inducer concentration, while recording their responses using fluorescence microscopy. The input perturbation sequence and calibrated single cell trajectories are then used to reconstruct the intracellular dynamics. Our Bayesian inference framework, Dynamic Prior Propagation (DPP) [259] allows the inference of unobserved molecular states and kinetic parameters, perform model selection tasks using Bayes factors, and dissect noise into intrinsic, extrinsic and technical components. Based on the inferred posterior distributions of kinetic parameters, a consecutive input perturbation can be selected, and the posterior distributions are used as prior distributions for the new iteration.

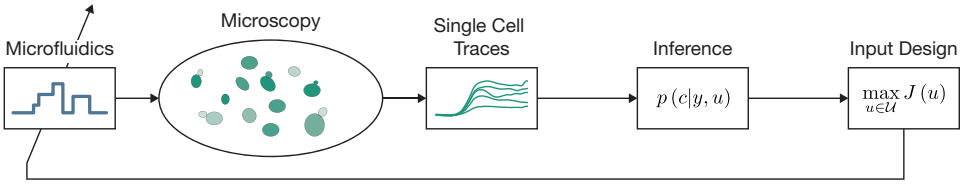

Figure 38: RECURSIVE BAYESIAN OPTIMAL EXPERIMENTAL DESIGN (OED) SCHEME. A microfluidic platform is used to synthesize perturbation sequences of inducer concentration during live cell fluorescence microscopy experiments. A recorded data set is then processed, mapped to molecule copy numbers and used to refine reaction rate estimates in the Bayesian Dynamic Prior Propagation (DPP) inference framework. Based on the resulting posterior distributions, a perturbation sequence for the next iteration can be designed.

A Microfluidic Platform for the Perturbation of Cellular Microenvironments

In order to synthesize temporal perturbation profiles in cellular microenvironments, we designed and manufactured a dedicated Polydimethylsiloxane (PDMS) microfluidic platform for the yeast *S. cerevisiae*, which combines our approach to Pulse Width Modulation (PWM) [6, 232] of liquid flows with single cell traps [46].

The chip design consists of several layers (see Figure 39 A). The *flow* part of the chip consists of two media inlets on the left, connected to reservoirs of cell medium (e.g. synthetic (SD) medium), and cell medium plus a maximum concentration of inducer concentration (e.g. SD medium + 0.5 M NaCl for inducing the High Osmolarity Glycerol (HOG) pathway), respectively. Hydrostatic pressure is evenly applied to both reservoirs to generate a constant media flow rate, equal for both inlets. The two media inlets converge into a mixing channel, connected to the cell chamber. The cell chamber hosts an array of 15 individual Field Of View (FOV) areas (Using a 60x objective and a Hamamatsu Orca Flash 4.0 camera. See section Materials & Methods.), each consisting of 67 cell traps, resulting in a theoretical maximum of > 1000 cells per experiment. An overlay of a bright field and CFP microscopy image of a single FOV can be seen in Figure 39 B, and shows the average loading efficiency of a single FOV of about 50 cells. Increasing the height of the cell chamber enhances the efficiency of removing cell clusters and capture individual single cells. However, this goes with decreasing imaging quality as cells might not stay in the focal plane.

On the right, the cell chamber is connected to a media outlet. At the top and bottom, two cell loading channels are connected to the cell chamber.

The *control* part of the chip consists of two microfluidic valves [233], that can be pressurized to squeeze the rounded media inlets below. By switching the two valves mutually exclusive, we can select one of the two input flows to be active and denote it as an *Off-State*, when cell medium is provided, and *On-State*, when the inducer medium is provided. Using the concept of PWM, we can define a short time period T_{PWM}, and denote the percentage of T_{PWM} in which the valves are in the *On-State* as the duty cycle D_{PWM}. By modulating D_{PWM}, we can synthesize medium concentrations between the two input flow concentrations. Figure 39 C shows an illustration of how modulating the duty cycle affects the output concentration. As the liquid pulses move through the mixing channel, the pulses slowly diffuse into each other. This *low-pass filter* is designed to filter the PWM switching frequency $f_{PWM} = 1/T_{PWM}$, but preserves the perturbation signal as $f_{PWM} << f_{\text{Perturbation Signal}}$. Figure 39 D shows two example profiles of fluorescent dye added to the input media, recorded by fluorescence microscopy. The left profile shows that arbitrary concentrations can be diluted, while the right profile shows that individual concentration levels can be kept constant over time.

Modeling Hog1 Induced Gene Expression

The transcriptional response to osmotic stress serves as a suitable model system to investigate the effect of various temporal perturbation sequences on the respective posterior distributions of kinetic rates.

We engineered a *Saccaromyces cerevisiae* strain, where the Yellow Fluorescent Protein (YFP) quadruple Venus (qV) is expressed under a STL1 Promoter (pSTL1). pSTL1 is an osmostress-inducible promoter, indirectly controlled through Hog1, the MAPK of the HOG pathway. Upon osmotic schock, Hog1 gets double phosphorylated and translocates into the nucleus, where it triggers a transcriptional response. While the increase of Hog1 activation upon increased NaCl concentration remains linear in a given range, the transcriptional response exhibits bimodal behavior [176]. Figure 40 A shows a dose response, measured by flow cytometry, of pSTL1-qV exression upon various concentrations of NaCl in the growth medium. While concentrations $\leqslant$ 0.2 M NaCl did not show any detectable expression of the reporter protein, levels above show two populations of expressing and non-expressing cells, with more cells shifting towards the active pool as the NaCl concentration increases.

In order to focus on a transcriptional model, the signaling part of the MAPK cascade was covered by an Ordinary Differential Equation (ODE) model, adapted from ref. [262]. Due to the inclusion of non-transcriptional and transcriptional feedback loops of glycerol production, the model can accurately predict the response to complex temporal input perturbations. Figure 40 B shows validation experiments comparing model predictions, using our refitted parameter values, to measurements of Hog1, fused to the Red Fluorescent Protein (RFP) mCherry, of the ratio

$$u(t) = \frac{\text{Hog1-mCherry}_{\text{Nucleus}}(t)}{\text{Hog1-mCherry}_{\text{Cell}}(t)},$$

which serves as the *input* for our transcription initiation model. Figure 40 C shows a schematic of the stochastic model with a repressed and a highly active transcriptional state [176]. All reactions are modeled according to mass action. The transition rate from the inactive to the repressed transcriptional state is modulated by the time-varying input $c_1 \cdot u(t)$. Extrinsic noise is captured by the translation efficiency c_8, drawn from a gamma distribution described by parameters α and β.

For an initial calibration of the transcription model, a simple step up in NaCl concentration, where we predicted the input sequence $u(t)$ with our signaling model and measured the qV abundance (see Figure 40 D), was used as a data set for a DPP run. The resulting posterior distributions over all kinetic parameters $(c_1, \ldots, c_7, c_9)$, the acquisition noise parameter (ω), and the extrinsic statistics (α, β) can be seen in Figure 40 E.

Refining Posterior Distributions

For a second DPP iteration, the posterior distributions from the initial calibration experiment (Figure 40 E) were used as prior distributions, with the extrinsic statistics (α, β) reset to initial values. To quantify the differences in information gain, Table 10 lists the KL divergence between the prior and posterior distributions for three different input sequences $u_j(t)$ (Figure 41 A). While even the single pulse experiment (i.e. input sequence $u_1(t)$) reduces parameter uncertainties during the second iteration, the most informative input sequence $u_2(t)$, which repeatedly activates Hog1, leads to a mean of the information gain increase over all kinetic rates of $\approx 43\,\%$, compared to the single pulse $u_1(t)$. Although the information gain for the significantly more complex input perturbation $u_3(t)$ is close to the one obtained using $u_2(t)$, it remains lower. Analysis of the single cell traces in Figure 41 B, showing the full dataset of input perturbation u_3, shows that most of the input dynamics is not reflected in the single cell recordings of qV expression. While this is most likely due to the slow reporter kinetics of expressing FPs, a more direct readout would be desirable.

	$u_1(t)$	$u_2(t)$	$u_3(t)$
c_1	1.40	1.88	1.79
c_2	1.46	2.05	1.71
c_3	1.54	2.13	2.00
c_4	1.50	2.18	1.96
c_5	1.43	2.06	1.82
c_6	1.66	2.01	1.91
c_7	1.54	2.50	1.94
c_9	1.52	2.37	2.12

Table 10: Kullback-Leibler divergence $D_{KL}\left[p\left(c_j \mid \mathbf{x}_u, u\right) \| \pi\left(c_j\right)\right]$ between prior and respective posterior distributions. The same prior distribution was used for the three DPP runs, each with a different input sequence $u_j(t)$.

Figure 41 C shows a comparison of the resulting posterior distributions of all kinetic parameters with the respective priors of input perturbation $u_3(t)$. A direct comparison of the posterior distributions, obtained using the simple step input $(u_1(t))$ and the input perturbation yielding the highest information gain $(u_2(t))$, can be seen in Figure 41 D.

DISCUSSION

We introduced a set of tools for optimizing the process of learning predictive computational models.

Our microfluidic device takes advantages from PWM in digital electronics to synthesizing temporal concentration profiles of inducer medium in

extracellular environments, as only two input flows of media are available and switched on and off, mutually exclusive. Capturing single cells within separated field-of-views allows unbiased recording using fluorescence microscopy, and simplifies the image processing tasks of segmenting and tracking cells.

While the demonstrated microfluidic platform, including the single cell traps, is intended for use with *S. cerevisiae*, a modified version can be readily applied to a variety of mammalian culture cells.

Our microfluidic platform allows to synthesize perturbation profiles that yield information rich datasets for the DPP inference framework. We chose the transcriptional response of pSTL1 to osmotic stress as a case study system and could show that experiments that repeatedly activate the MAPK Hog1, can reduce parameter uncertainties significantly better than sequences of lower complexity. The slow dynamics of fluorescent reporter proteins as expression reporters, however, remains a limiting step. The application of direct readouts of mRNA production, for instance using the MS2 or PP7 systems [129, 127], could be a potential solution to follow the transcriptional dynamics closer.

Based on available posterior distributions, we could now simulate cell trajectories, test proposed input sequences *in silico* for their expected information gain, and validate the predictions experimentally. In a simulation case study, we already showed how to design perturbation sequences by formulating a variational problem and solving it using a stochastic approximation algorithm [256]. Due to the computational complexity involved, the optimization of temporal perturbation sequences is currently an ongoing task.

MATERIALS & METHODS

Strains & Plasmids

Yeast strains and Plasmids are listed in Tables 11 and 12.

Flow Cytometry

Single colonies of the yeast strain yMU53 were inoculated in synthetic (SD) medium and grown overnight at 30° C. Saturated cultures were diluted and grown in log phase for at least two doubling times (> 4 h), before the HOG pathway was induced by adding 100 µl of SD medium plus the respective concentration of NaCl, to 200 µl cell suspension, such that the intended final NaCl concentration was reached. Translation was blocked after 45 min by adding 100 µl of Cyclohexamide (CHX) to a final concentration of 0.1 mg ml^{-1}. After incubation for 1.5 h the cells were measured on a flow cytometer (FACSCalibur™, BD).

Data was analyzed and plotted using custom MATLAB (The MathWorks) scripts. Acquired data was gated to retain single cell measurements.

Fluorescence Microscopy

Experiments were performed on a fully automated, inverted epifluorescence microscope (Eclipse Ti, Nikon Instruments), 60x (NA 1.4) oil immersion objective, specific (CFP/YFP/mCherry) excitation and emmission filters, and illuminated using a SPECTRA X light engine (Lumencor). The microscope was located in an incubation chamber, set to maintain a constant temperature of 30° C. Imaging settings and parameters were kept constant for all experiments. Multiple field-of-views were recorded for each time point, using a motorized xy stage, while the focal plane was maintained using a Perfect Focus System (PFS) (Nikon Instruments). The microscope and related components were computer controlled using μManager [58].

Single colonies of the respective yeast strain were inoculated in synthetic (SD) medium and grown overnight at 30° C. Saturated cultures were diluted and grown in log phase for at least two doubling times (> 4 h). The cell cultures were diluted again before loading into microfluidic chips (OD600 = 0.05).

Image Analysis

Microscopy images were analyzed using the YeastQuant [175] platform. For images of the strain yMU53, the HTA2-CFP nuclear marker was used to segment the nucleus and locate cells. Cell boundaries were detected as secondary objects surrounding the nucleus in the unbiased bright-field images. Strains without a nuclear marker were segmented on the basis of two bright-field images, and the total cellular qV intensity was calculated. Illumination conditions were kept constant for all experiments.

Microfluidics

Custom PDMS chip designs were drawn using AutoCAD (Autodesk) and individual layers were printed on to chrome on glass masks (Front Range Photomask). Wafer molds were produced in house (For a detailed description of the wafer manufacturing process and the photoresists, see Section A.1). Wafer molds were treated with 1H,1H,2H,2H-Perfluorodecyltrichlorosilane (abcr) before PDMS was first applied. The *control layer* was fabricated using a ratio of 5:1 of Sylgard 184 (Dow Corning) and curing agent. The *flow layer* was fabricated using a 20:1 ratio. (A detailed protocol of the chip fabrication can be found in Section A) After the fabrication, inlet and outlet holes were punched, and the PDMS devices bonded to glass coverslips using UV treatment.

Inlets were connected to hydrostatic pressure driven reservoirs using PTFE tubing. Valves were pressurized using three-way solenoid valves (The Lee Company), computer controlled using an USB controller board (National Instruments) and custom MATLAB (The MathWorks) scripts.

Quantification of Protein Levels

We applied a mapping of fluorescence intensities, recorded by microscopy experiments, to absolute copy numbers of fluorescent proteins, based on one we established in a previous study [259]. To increase the mapping precision, we selected more proteins with homogeneous distribution in the cytoplasm [104] and approximate copy numbers [75] up to $\approx$ 50000. Selected proteins are listed in Table 13. A detailed description of the quantification method can be found in Supplementary Note 4 of Chapter 6.

Inference Algorithm

We applied the Bayesian inference framework of Dynamic Prior Propagation for the inference of stochastic biochemical processes from pooled single cell time lapse measurements as introduced in ref. [259]. A detailed description can be found in Capter 6.

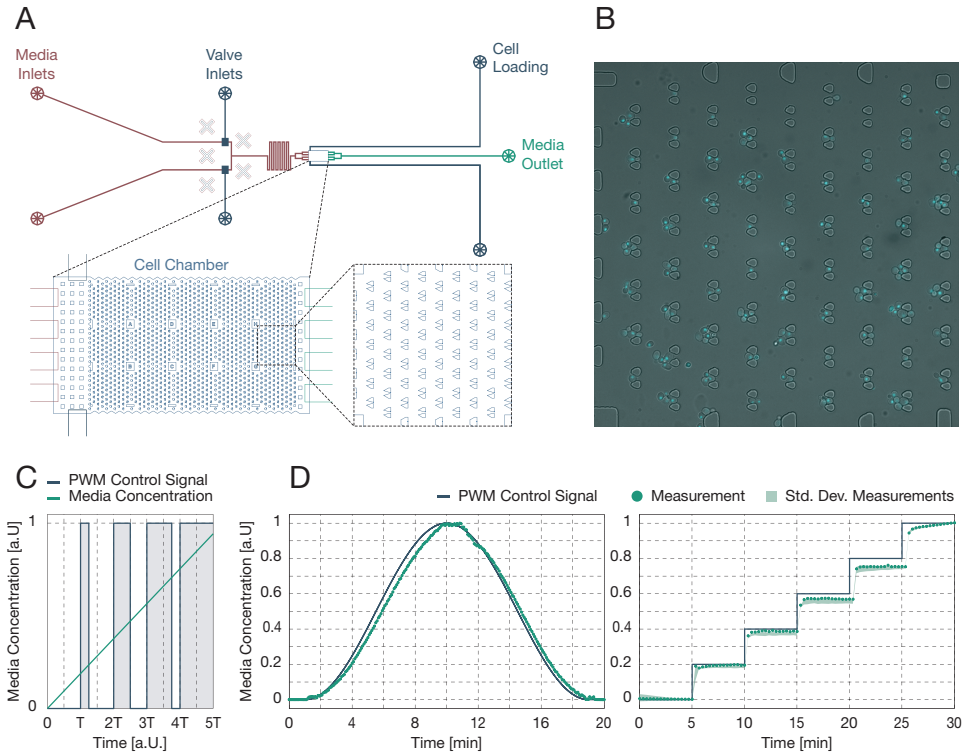

Figure 39: MICROFLUIDIC PLATFORM TO SYNTHESIZE TEMPORAL PERTURBATION PROFILES. (A) Drawing of the multi-layer (separated by individual colors) microfluidic PDMS valve-on-chip platform that combines the synthesis of temporal perturbation profiles using Pulse Width Modulation (PWM) of liquid flows and the capturing of single cells using individual traps. The dashed lines mark zoomed out regions including the imaging chamber, and an individual field of view (Using a 60x objective and a Hamamatsu Orca Flash 4.0 camera. See section Materials & Methods.). (B) Overlay of bright-field and CFP (HTA2-CFP nuclear marker) microscopy images showing captured cells in a single field of view. (C) PWM schematic illustrating how switching between the two input flows in a specific ratio affects the media concentration in the cell chamber. (D) Recordings of temporal profiles using fluorescent dye illustrating how (left) arbitrary values between the two input concentrations can be diluted, and how (right) individual concentration values can be kept constant over time.

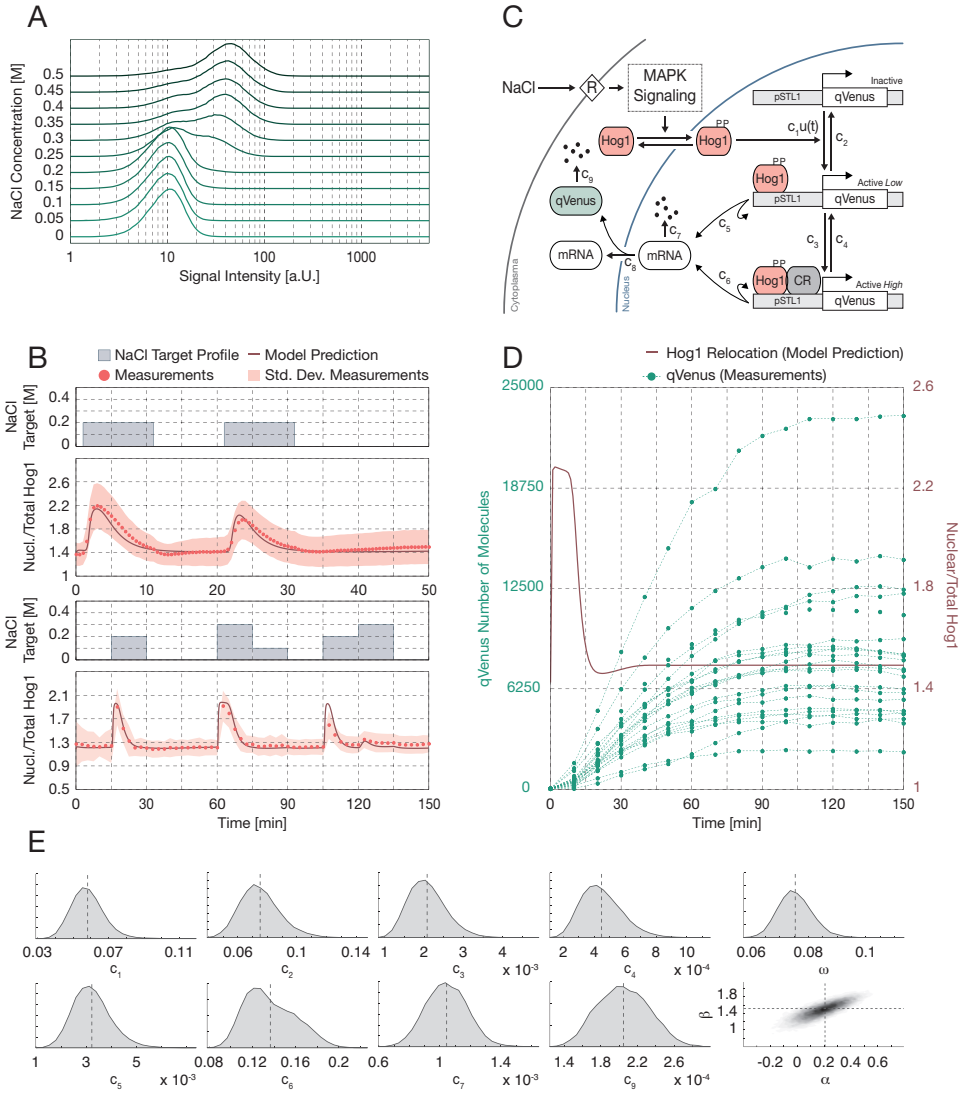

Figure 40: MODELING HOG1 INDUCED GENE EXPRESSION. (A) Dose response of a quadruple-Venus fluorescent reporter under the STL1 promoter, measured by flow cytometry. For intermediate levels of NaCl concentration, a bimodal expression pattern can be observed on the population level. (B) Validation of an ODE model of the MAPK signaling cascade to map temporal profiles of NaCl concentration to Hog1 relocation. (C) Stochastic mass-action model of Hog1 induced gene expression. Extrinsic noise is introduced by a variable translation efficiency c_8. (D) Initial dataset of qV abundance upon NaCl addition. Note that the time delay between Hog1 relocation and qV expression was removed for visualization, and as protein maturation steps are not explicitly modeled (see also Chapter 6, Supplementary Note 4). (E) DPP was initially performed with data from panel (D) using 50,000 samples. The posterior distributions over all kinetic parameters ($c_1, \ldots, c_7, c_9$), the acquisition noise parameter (ω), and the extrinsic statistics (α, β) are shown.

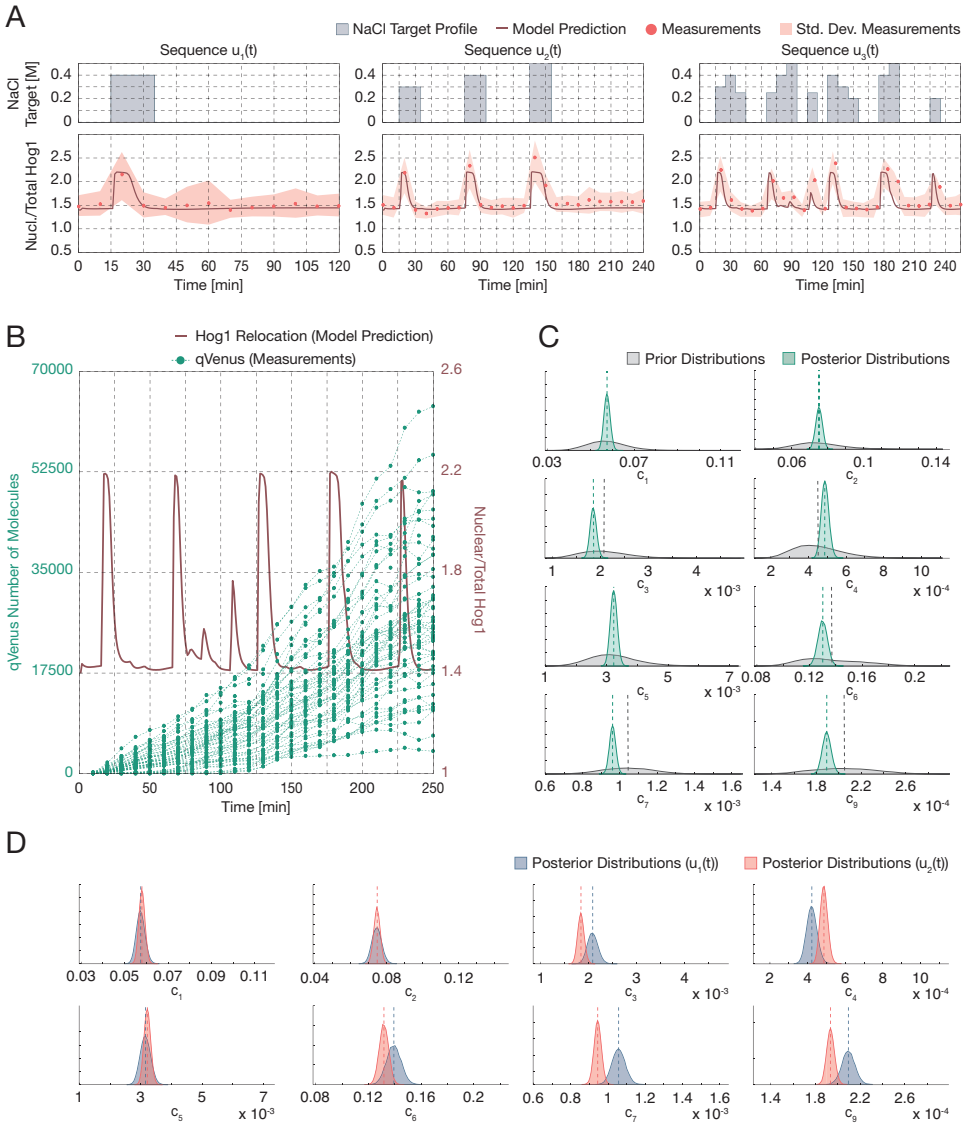

Figure 41: REFINING POSTERIOR DISTRIBUTIONS. (A) Various input sequences $u_j(t)$ used to measure cellular responses and calibrate the transcriptional model. (B) Complete dataset of a temporal perturbation sequence $u_3(t)$ of NaCl concentration, including the signaling model prediction and measured qV abundance, used for the second DPP iteration. Note that the time delay between Hog1 relocation and qV expression was removed for visualization, and as protein maturation steps are not explicitly modeled (see also Chapter 6, Supplementary Note 4). (C) Comparison of prior and posterior distributions for the second DPP iteration, using input perturbation sequence $u_3(t)$. (D) Direct comparison of the posterior distributions, obtained using the simple step input ($u_1(t)$) and the input perturbation yielding the highest information gain ($u_2(t)$).

.

STRAIN ID	GENOTYPE	SOURCE
yBH66	BY4741 *(his3Δ1; LEU22Δ0; met15Δ0; URA33Δ0; MATa)*	Open Biosystems
yMU53	*HTA2-CFP; Hog1-mCherry::URA3; STL1-qV::LEU2*	This study
yMU55	OAF3-qV::URA3	This study
yMU56	FMP48-qV::URA3	This study
yMU57	FRK1-qV::URA3	This study
yMU58	SYH1-qV::URA3	This study
yMU59	TMT1-qV::URA3	This study
yMU60	YGF117C-qV::URA3	This study
yMU61	YHR112C-qV::URA3	This study
yMU62	GPX2-qV::URA3	This study
yMU63	MET14-qV::URA3	This study
yMU64	REH1-qV::URA3	This study
yMU65	TMA46-qV::URA3	This study
yMU66	TMA108-qV::URA3	This study
yMU67	OCA4-qV::URA3	This study
yMU68	SSE2-qV::URA3	This study
yMU69	HOG1-qV::URA3	This study
yMU70	TRM732-qV::URA3	This study
yMU71	FRD1-qV::URA3	This study
yMU72	TDA1-qV::URA3	This study
yMU73	APT1-qV::URA3	This study
yMU74	YGR210C-qV::URA3	This study
yMU75	OCA5-qV::URA3	This study
yMU76	NPA3-qV::URA3	This study
yMU77	CPA2-qV::URA3	This study
yMU78	YNL247W-qV::URA3	This study
yMU79	HSP104-qV::URA3	This study
yMU80	OLA1-qV::URA3	This study
yMU81	CAR1-qV::URA3	This study
yMU82	GUS1-qV::URA3	This study

Table 11: Yeast Strains

PLASMID ID	GENOTYPE	SOURCE
pMU7	TAPhom-2HA-qV-URA3	this study
pSP34	pSTL1-qV-LEU2	S. Pelet

Table 12: Plasmids

STANDARD NAME	SYSTEMATIC NAME	MOLECULES PER CELL
YKR064	OAF3	149
YGR052W	FMP48	339
YPL141C	FRK1	556
YPR105C	SYH1	830
YER175C	TMT1	937
YGR117C	YGR117C	1280
YHR112C	YHR112C	1360
YBR244W	GPX2	2010
YKL001C	MET14	2170
YLR387C	REH1	2240
YOR091W	TMA46	4220
YIL137C	TMA108	5110
YCR095C	OCA4	5370
YBR169C	SSE2	6300
YLR113W	HOG1	6780
YMR259C	TRM732	7110
YEL047C	FRD1	7620
YMR291W	TDA1	10200
YML022W	APT1	11200
YGR210C	YGR210C	11600
YHL029	OCA5	12500
YJR072C	NPA3	15200
YJR109C	CPA2	18000
YNL247W	YNL247W	23000
YLL026W	HSP104	32800
YBR025C	OLA1	36800
YPL111W	CAR1	42800
YGL245W	GUS1	48700

Table 13: List of Reference Proteins

Part III

CELL-CELL COMMUNICATION: INVESTIGATING THE SENSING AND DECODING OF SHALLOW CHEMICAL GRADIENTS

8

MANUSCRIPT 6

REFERENCE

B. Hegemann, M. Unger, S.S. Lee, I. Stoffel-Studer, J. van den Heuvel, S. Pelet, H. Koeppl, and M. Peter, *A Cellular System for Spatial Signal Decoding*, Submitted

AUTHOR CONTRIBUTIONS

BH, MU, SSL, IS and JvH performed experiments. BH, MU, HK, SP and MP participated in experimental design and model construction. BH, MU and MP wrote the paper.

A CELLULAR SYSTEM FOR SPATIAL SIGNAL DECODING

B. Hegemann[1,5], M. Unger[1,2], S.S. Lee[1], I. Stoffel-Studer[1], J. van den Heuvel[1], S. Pelet[3], H. Koeppl[4], and M. Peter[1,5]

ABSTRACT

Cell-cell communication requires cells to navigate along chemical gradients, but how the gradient directional information is identified remained elusive. We use single cell analysis and mathematical modeling to define the cellular gradient decoding network in yeast. Our results demonstrate that the spatial information of the gradient signal is read using double positive feedback between the GTPase Cdc42 and trafficking of the receptor Ste2. Spatial decoding critically depends on low Cdc42 activity which is maintained by the MAPK Fus3 through sequestration of the Cdc42 activator Cdc24. Deregulated Cdc42 or Ste2 trafficking prevents gradient decoding and leads to mis-oriented growth. Our work discovers how a conserved set of components assembles a network integrating signal intensity and directionality to decode the spatial information contained in chemical gradients.

ONE SENTENCE SUMMARY

Our work identifies the molecular details of a system for spatial signal decoding in yeast and derives general design principles for gradient sensing systems.

INTRODUCTION

Directional information encoded in chemical gradients is required for numerous processes such as amoeba and neutrophil chemotaxis or chemotropic growth in yeast and neurons (Fig. 42 A). Although very diverse, all gradient sensing systems use a similar set of cellular pathway components to solve the task of navigating cells towards the gradient source. Each contributing cellular pathway is understood at considerable detail and interaction within and between pathways is predominantly believed to be hierarchical [88]. However, linear pathway organization has been insufficient to explain the emergent property of gradient decoding for successful cell navigation.

1 Institute of Biochemistry, D-BIOL, ETH Zurich, Zurich, Switzerland
2 Automatic Control Laboratory, D-ITET, ETH Zurich, Zurich, Switzerland
3 Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland
4 Department of Electrical Engineering and Information Technology, Darmstadt University of Technology, Darmstadt, Germany
5 Correspondence to: bjoern.hegemann@bc.biol.ethz.ch (B.H.) and matthias.peter@bc.biol.ethz.ch (M.P.)

Gradient decoding can be formalized by dividing the gradient signal into two: an intensity signal and a directional signal. Intensity is decoded by signal integration across the cell surface mediated by the signal receptors and downstream signaling pathway (reviewed in [40]). Direction decoding is believed to require computing the signal difference between cell front and back in gradients (Fig. 42 B [199]). Although the front-back difference is conceptually simple, experimental identification of the network component computing this global front-back difference has not been successful. Front identification is thought to occur prior to activation of downstream polarity pathways which are centered on GTPases initiating cytoskeletal rearrangements. GTPase systems can auto amplify and establish polarity even in the absence of spatial cues (e.g. during symmetry breaking in cultured neurons [52] or yeast [240]) and are thus expected to initiate polarized growth in random directions if activated precociously. Although the polarized growth direction is adjustable [56], in vivo observations report high directional accuracy of initial symmetry breaking towards chemical gradients [5, 107] and thus suggest that gradients are decoded before polarity pathways become activated.

Since a conserved set of components in diverse systems can decode shallow gradients [24], we chose the prototypic gradient sensing system of the budding yeast pheromone response to dissect how these components assemble into a network to localize the gradient source. We applied a combination of quantitative microscopy, microfluidic gradient generators and computational modelling to study the dynamics of polarity established towards a defined chemical gradient. This approach allowed us to define and dissect the core network required for gradient sensing and to quantify the contribution of its key nodes. We establish that trafficking of the receptor and position of the polarity site are connected in a spatial double positive feedback loop to decode the directional gradient signal locally. Precise control of Cdc42 GTPase activity is crucial for network function and depends on the MAPK Fus3 that controls intensity signal-dependent sequestration of the Cdc42 activator Cdc24. The identified network is based on conserved components and thus might form the core of many gradient-sensing systems.

RESULTS

Gradient sensing is directed by a mobile polarity complex

To identify subcellular localization dynamics of different components during the gradient response, we developed a set of live cell microscopy assays using microfluidic gradients [132]. Since cells are only susceptible to α-factor during the G1 phase of the cell cycle [41], we defined G1 entry as time zero of the gradient response. The GTPase Exchange Factor (GEF) Cdc24 activates the polarity GTPase Cdc42 and thus serves as the earliest reporter for polarity establishment (Fig. 42 C). In a microfluidic gradient

Cdc24 fused to quadruple Venus (qV) relocated mainly to the nucleus and at the same time established polarity on the plasma membrane within the first minutes of gradient exposure (Fig. 42 D and Movie S1). Surprisingly, the established polarity axis was not aligned with the external gradient. Rather, the polarity site followed several intensity fluctuation cycles to adjust the polarity axis towards the gradient before polarized growth was initiated. We observed identical behavior in natural gradients where similar cycles aligned the polarity axis with the partner cell before cell-cell fusion was initiated (Fig. 46 A and Movie S2). Importantly, this tracking behavior was also detectable for the Cdc42 effector Bni1 (Fig. 42 E), thus suggesting oscillatory activation of the entire polarity pathway. Quantification across cell populations confirmed that polarity axis direction was largely gradient direction independent at the time of polarity site establishment (t_{PE}) while at onset of polarized growth (t_{PG}) the axis was aligned with the gradient (Fig. 42 F). Cells that established polarity further away from the gradient took longer to align their polarity axis and initiate polarized growth (Fig. 42 G), suggesting that site assembly is biased towards the higher gradient signal and moves the polarity axis in small steps along the membrane. Based on these results, we concluded that the site of polarity establishment is independent of the gradient direction and that gradient sensing may require a mobile polarity site (Fig. 42 H).

Intensity signal regulates Cdc42 activity and polarity site mobility

If polarity site mobility serves to scan for a higher gradient signal, we reasoned that it should depend on coupling the polarity site to the activated receptor. We generated a mutant in the polarity scaffold Far1, which is unable to interact with the activated G$\beta\gamma$ dimer (Far1-ΔG$\beta\gamma$ bind [41]). Cells expressing this mutant still displayed polarity site mobility when exposed to an α-factor gradient, but were unable to align the site with the gradient and initiated polarized growth in random directions (Fig. 47 A, B, Movie S3). When we exposed wild-type cells to a uniform field of α-factor, they still exhibited a highly mobile polarity site (Fig. 43 A and Movie S4). However, exposing cells to a uniform field of high α-factor strongly reduced site mobility (Fig. 43 B), leading to much earlier t_{PG} compared to low uniform α-factor or gradients (Fig. 43 C). Importantly, polarized growth in high α-factor was not initiated in random directions with respect to the position at t_{PE} but instead proximal to the site of polarity establishment (Fig. 47 C), further confirming reduced site mobility. These observations show that the components decoding the gradient intensity signal, i.e. receptor and Fus3 signaling, regulate polarity site mobility independent of a gradient (Fig. 43 D). To test whether polarity site mobility arises from partial Cdc42 activation and an unstable polarity complex, we quantified the levels of the Cdc42 effector Bni1 in different α-factor concentrations. Cdc42 activity increased with increasing α-factor concentrations (Fig. 43 E) and thus confirmed that the intensity signal regulates polarity complex activa-

tion and suggest that polarity site mobility depends on Cdc42 activation levels.

Fus3 driven Cdc24 sequestration regulates Cdc42 activity

Cdc42 activation is a self-amplifying process fed by two positive feedback loops, an immediate one through further recruitment of Cdc24 by Bem1 [41, 35] and a delayed loop based on actin cable-mediated Cdc42 trafficking [241]. How activity of these non-linear processes is regulated remains unclear. Inspired from previous theoretical work [112, 86] and the fact that the Cdc42 activator, Cdc24, is mainly nuclear [164, 205], we hypothesized that Cdc24 sequestration may limit the Bem1 based Cdc42 activation loop. We performed dose response experiments to test if α-factor concentration regulates nuclear Cdc24 levels. We confirmed that Cdc24 is mostly nuclear upon G1 phase entry and further found that Cdc24 nuclear levels were negatively correlated with α-factor concentration (Fig. 43 F). Concurrently, membrane-bound Cdc24 increased with higher α-factor (Fig. 43 G) while total Cdc24 levels remained constant (Fig. 48 A). α-factor-induced Cdc24 relocation was dependent on Fus3 since inhibition of Fus3 fully reversed α-induced Cdc24 nuclear export (Fig. 48 B). We concluded that Fus3 regulates cytosolic availability of the Cdc42 activator Cdc24.

In a second set of experiments we tested how Fus3 regulates Cdc24 cytosolic levels. Cdc24 nuclear localization depends on Far1 [164, 205], and Far1 expression is induced with increasing α-factor (Fig. 48 C). Interestingly, Far1 was exported from the nucleus in an α-factor-dependent manner (Fig. 48 D [27]), and acute inhibition of Fus3 immediately decreased Cdc24 cytosolic levels while concurrently reducing membrane-bound Cdc24 (Fig. 48 E and 43 H). Far1 is phosphorylated by Fus3 on S341 and S346 [72] located within the Far1 Nuclear Export Signal (NES) (Figure 48 F [27]). Deletion of the Far1 export sequence (Far1-ΔNES) or mutation of S341 and S346 to non-phosphorylatable alanine residues (Far1-NES2A) interfered with Fus3-dependent Cdc24 export resulting in decreased cytoplasmic accumulation of Cdc24 (Fig. 48 G and 48 H). Conversely, partial mutation of the Far1 nuclear localization sequence (Far1-pNLS) increased cytosolic Cdc24 levels while Fus3-induced Cdc24 nuclear export remained intact (Fig. 48 I). Together, these data demonstrate that Fus3 directly controls Cdc24 nuclear export through Far1-NES phosphorylation and drives Cdc24 nuclear import through increasing Far1 expression.

In a third set of experiments we addressed whether Cdc24 cytosolic levels alone regulate Cdc42 activity. Deletion of the Far1 NES or the Far1 NLS (Far1-ΔNLS) decreased or increased Cdc24 cytosolic levels, respectively (Fig. 48 G and 48 J). We thus quantified Cdc24 membrane levels in these two mutants and found that low cytosolic Cdc24 in Far1-ΔNES prevented increased Cdc24 levels at the membrane (Fig. 43 I), while excess cytosolic Cdc24 in Far1-ΔNLS increased Cdc24 membrane levels (Fig. 43 J). As a con-

sequence, polarity site mobility was decreased with excess cytosolic Cdc24 even at low α-factor, and in high α-factor conditions polarity site mobility remained high when Cdc24 nuclear export was prevented (Fig. 43 K). In summary, these experiments identify a regulatory network for how the gradient intensity signal regulates polarity site mobility (Fig. 43 L): Cytosolic Cdc24 availability drives membrane-bound Cdc24, thus Cdc42 activity and polarity site mobility.

Tight control of cytosolic Cdc24 levels is achieved using a regulatory system resembling an Incoherent Feed Forward Loop (IFFL). An IFFL consists of a single component, Fus3 in the present case, driving both negative and positive regulatory loops [93]. Fus3 negatively regulates cytosolic Cdc24 levels with a delay by inducing Far1 expression leading to Cdc24 nuclear sequestration. Concurrently Fus3 positively regulates cytosolic Cdc24 on a faster time scale by phosphorylation-dependent Far1-Cdc24 nuclear export.

Double positive feedback network topology for gradient decoding

Our initial observations of a mobile polarity site suggested that polarity site mobility plays an active part in gradient decoding. To identify how a mobile polarity site decodes the gradient, we developed an exploratory computational model that incorporated our experimental results into a generalizable network of conserved components. The model is centered on the GTPase activation loop whose activity depends solely on GEF cytosolic availability (Fig. 44 A, reactions 3-6 [112]). To direct the GTPase complex towards the gradient, we connected the polarity complex (N) to the gradient receptor (R) by recruiting cytosolic N (Nc) to active membrane-bound receptor (Rm) and by making part of R exo- and endocytosis spatially dependent on membrane bound N (Nm, Fig. 44 A, reactions 3 & 4 [34, 163] and 6 & 7 [14], respectively; see Supplementary Materials and Fig. 49 A-E for details). This simplified model resembles a double positive feedback loop capable of transforming graded input into switch-like output [65, 31]. Here, the graded input is the gradient signal which linearly increases along the membrane and switch-like output is the polarity site which focuses all GTPase activity within a small membrane region. We tested this model and found that the polarity site formed by Nm fluctuated and adjusted gradually towards the gradient source, much like our experimental results. Moreover, the direction at t_{PG} (see Supplementary Materials for how we determine t_{PG} in the model) across many simulations resembled our in vivo data from microfluidic gradients (Fig. 44 B), demonstrating that we could use the model to dissect how the network components are involved in gradient decoding.

Is limitation of Nc that is available for auto amplification of Nm sufficient for an adjustable polarity axis? Simulations run without a gradient did result in a site that dis- and re-assembled along the membrane similar to experimental data (Fig. 44 D). Next we tested how tightly N levels

need to be controlled to ensure an adjustable polarity axis. Low levels of N yielded a fluctuating polarity site, however the site failed to stabilize towards the gradient such that t_{PG} was never reached. Conversely, high levels of N yielded a very stable polarity site that did not translocate along the membrane and consequently reached a position at t_{PG} that was independent of the gradient direction (Fig. 44 E and F). These simulations thus support our experimental data and show that activator limitation is sufficient for controlling GTPase activity and for maintaining an adjustable polarity axis. In addition, controlled GTPase activity is required for gradient decoding.

How is limited GTPase activity important for gradient decoding? The number of Nm, i.e. active GTPases, directly correlates with polarized receptor trafficking rates and thus determines the ratio between polarized (Nm-dependent) and unpolarized (Nm-independent) receptor trafficking (ratio between reactions 6 & 7 and 1 & 2, respectively (Fig. 44 A and 49 A)). Overall rate changes did not affect gradient sensitivity (Fig. 49 F) and neither did decreasing the polarized traffic to levels 10-fold lower than unpolarized traffic. However, when polarized traffic was 40-fold higher than unpolarized traffic, resembling a state of polarized growth [146], faithful gradient decoding was prevented (Fig. 44 G). Together, the theoretical approach predicts that limited GTPase activation serves two functions: It prevents polarity site stabilization and thus permits continuous polarity axis adjustment towards the gradient. In addition, limited GTPase activation decreases polarized receptor traffic and as such enables spatial signal decoding by the polarity and trafficking pathways feedback.

Key quantitative links for building the gradient decoding network

To test whether reduced receptor trafficking and limited GTPase activity are indeed critical for gradient sensing, we designed experiments to observe and manipulate these two network components in vivo in the yeast system. In control cells, general exocytosis and endocytosis markers (Exo70 and Abp1, respectively) as well as the receptor Ste2 translocated along the membrane, and had lower or a more distributed intensity during the gradient decoding phase. Intensity was increased and more focused upon initiation of polarized growth (Fig. 45 A, B, 50 A and Movie S5-7). To test if limited trafficking was sufficient for these observed dynamics, we reduced polarized trafficking by inhibiting the actin cytoskeleton with Latrunculin A (LatA). Importantly, we found that the Cdc24 polarity site was as mobile as in control cells (Fig. 45 C, 50 B, C and Movie S8), and likewise the behavior of Exo70 and Ste2 exocytosis during the gradient decoding phase was independent of the actin cytoskeleton (Fig. 45 C, 50 D, E and Movie S9 and S10). To test if increased polarized traffic prevents gradient sensing, we removed limitation of polarized traffic using the Far1-ΔNLS mutant (Fig. 43 J) and followed Cdc24 in cells exposed to α-factor gradients. We found a less mobile polarity site that largely impaired gradient

decoding (Fig. 45 D, E and Movie S11). These results establish that in vivo polarity and polarized trafficking is limited during gradient decoding and that this limitation is required for a sufficiently mobile polarity site and gradient decoding.

Limitation of polarized traffic and Cdc42 activation is achieved by the inhibitory arm of the Fus3-driven IFFL through nuclear sequestration of Cdc24. What is the role of the activating arm, namely Fus3 driven Cdc24 nuclear export? We deleted the Far1 NES and followed Cdc24 in α-factor gradients. These cells still displayed a mobile polarity site which stabilized at random positions on the membrane and initiated polarized growth independent of the gradient direction (Fig. 45 F, G and Movie S12). This experiment establishes that further polarity activation is required for site stabilization once it has aligned with the gradient. By linking receptor activation through Fus3 to the level of polarity activation the identified IFFL thus serves as the master regulator for gradient decoding.

DISCUSSION

Directed movement along chemical gradients is an evolutionarily conserved process essential for single and multicellular eukaryotes. Yet the mechanisms identifying the gradient directional information remain poorly understood. Here we used chemical and genetic perturbation experiments combined with computational modeling to study gradient decoding in yeast. Our results define how cells integrate the directional signal, using Cdc42 to Ste2 spatial feedback, and the intensity signal, using Fus3 controlled Cdc42 activity, to process and interpret the spatial signal contained in chemical gradients (Fig. 45 H).

A core network topology for spatial signal processing

Cell polarity is established by two positive feedback loops, in which an immediate loop breaks cell symmetry (GTPase activation [28, 35]) and a second, delayed loop stabilizes the polarity axis position (actin-based GTPase membrane delivery [241]). Here we find that during initial gradient sensing, actin-based feedback remains inactive and is replaced by receptor trafficking feedback. Receptor trafficking stabilizes the polarity axis position dependent on the external gradient signal, thus aligning polarity with the gradient. A feedback linking the gradient sensor and downstream polarity may present a generalizable concept for gradient amplification, especially in systems in which the gradient receptor becomes polarized such as in chemotropic neuronal polarization [5]. In other systems, such as chemotaxing cells, the activated G$\beta\gamma$ complex and not the receptor is polarized [250, 169]. Since recent evidence suggests that Rho GTPases are required for gradient sensing independent of the actin cytoskeleton in chemotaxing cells [239], it is conceivable that GTPase-directed trafficking

of the G$\beta\gamma$ complex constitutes the spatial double positive feedback loop in chemotaxis.

Localized decisions and time averaging eliminate noise

The molecular basis of the central concept in gradient decoding, calculation of the front to back signal difference, has long remained elusive. Here we identify how linking receptor trafficking and polarity localizes the difference calculation to a focused membrane region (Fig. 45 I). Within this region cycles of polarity site dis- and reassembly are biased towards higher active receptor density. Biased polarity site re-assembly thus establishes a succession of localized directional decisions adjusting the cellular polarity axis towards the gradient source in a biased random walk. Repeated localized decisions may include *wrong* directional decisions, which are cancelled out by time-averaging. Once aligned with the gradient, polarity site reassembly becomes spatially confined, leading to further polarity amplification by actin- and receptor-mediated feedback, eventually initiating polarized growth. This system thus relies on molecular dynamics and time averaging across repeated directional decisions to decode the noisy gradient signal.

Incoherent feed-forward control of non-linear systems

The spatial feedback system identified here critically depends on tight control of Cdc42 activity. GTPase self-amplifying signaling cascades can establish polarity even in the absence of spatial cues [52, 240, 105], functioning as molecular switches. Here we identify partially activated Cdc42 as a stable state controlled by sequestration of the Cdc42 GEF, Cdc24. We find that a Fus3-driven incoherent feed forward loop constitutes a robust regulatory mechanism for Cdc24 sequestration and thus Cdc42 activity. Incoherent feed forward loops have been identified in various gene regulatory networks as a robust and time integrating system [93]. Here, the slow inhibitor arm of the network buffers the system against signaling noise while the fast activator arm allows fine-tuning of the output. A similar incoherent feed forward loop downstream of the GTPase Rac was recently identified in chemotaxis [48]. Analogous to yeast Fus3, Rac drives actin polarization through Arp2/3 activation while at the same time recruiting the Arp2/3 inhibitor Arpin. Removal of the inhibitor increases polarization activity while preventing cell steering during chemotaxis, in striking resemblance to our data observed with cells where Cdc42 limitation was removed. These cells were defective in gradient sensing due to premature onset of polarized growth. Spatial exclusion of an activator under control of a robust feed forward loop may thus present a generalizable mechanism allowing regulation of non-linear processes such as GTPase activation.

Taken together, our work provides the foundations for understanding how distinct signaling pathways assemble a regulatory network capable of decoding spatial signals.

ACKNOWLEDGMENTS

We thank S. Altschuler (UCSF, US) for sharing his model script, J. Ouellet (ETH Zurich, CH) for the GFP replacement plasmid, members of the Peter laboratory for helpful discussions, and R. Dechant, B. Fierz, S. Märki and A. Smith for critical reading of the manuscript. BH was funded by a DFG Fellowship, MU by a joint PhD fellowship by SystemsX.ch, and SP by a Förderprofessorship of the Swiss National Science Foundation (SNF). Work in the Koeppl laboratory is supported by the SNF, and work in the Peter laboratory by an ERC senior award, the SNF and ETHZ. The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

BH, MU, SSL, IS and JvH performed experiments. BH, MU, HK, SP and MP participated in experimental design and model construction. BH, MU and MP wrote the paper.

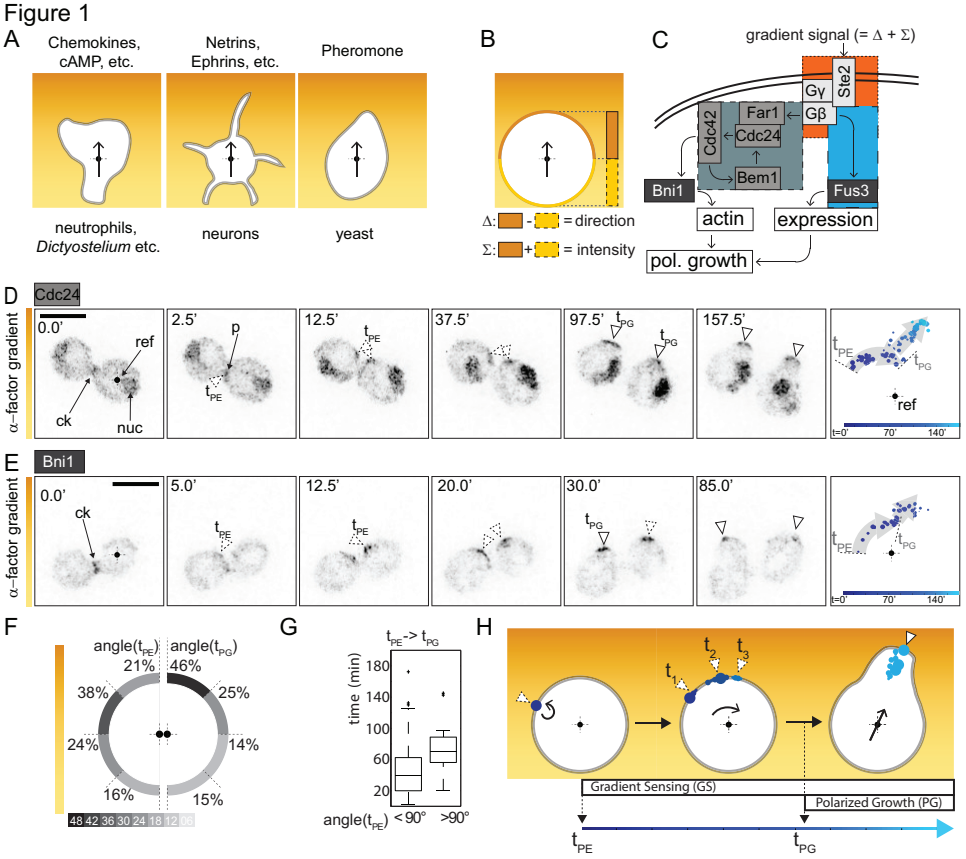

Figure 42: *Gradient sensing is directed by a scanning polarity site.* (A) Shallow chemical gradients guide cells towards their targets. (B) The gradient signal transmits two type of information. (C) Scheme of yeast mating pathways: the polarity pathway (GTPase Cdc42, GEF Cdc24, scaffolds Far1 and Bem1 (dark grey shading)) and its downstream effectors (Bni1, actin polymerization), the receptor pathway (GPCR Ste2, activated G-proteins Gβ and Gγ (orange shading)) and the simplified MAPK pathway (MAPK Fus3 (blue shading)). (D) Cells expressing Cdc24-qV exposed to a microfluidic α-factor gradient resulting in $a \approx 2.5\,\mathrm{nM}$ concentration difference across the cell. Labeled are time in minutes, cytokinesis position (ck), Cdc24 nuclear (nuc) and polarity site (p) localization. Dashed arrow heads for polarity establishment (t_{PE}) and scanning site, closed arrow heads for onset (t_{PG}) and continuing polarized growth. Last panel represents the time projected position of the polarity site relative to cell center (ref): site position, intensity (circle size) and time (color). Scale bar = 5 µm. (E) Cells expressing Bni1-qV were treated as in (D). (F) Angle of polarity site with respect to gradient at indicated time points was determined from experiment in (D). Results binned in 45° increments and expressed as a percentage (experiments (N) = 3, average number of cells ($n_{Ø}$) = 86). (G) Extent of polarity site scanning (time from t_{PE} to t_{PG}) determined from experiment in (D) for cells establishing polarity more or less than 90° away from the gradient (N = 3, $n_{Ø}$ = 48). (H) Schematic of polarity complex dynamics during gradient sensing.

Figure 2

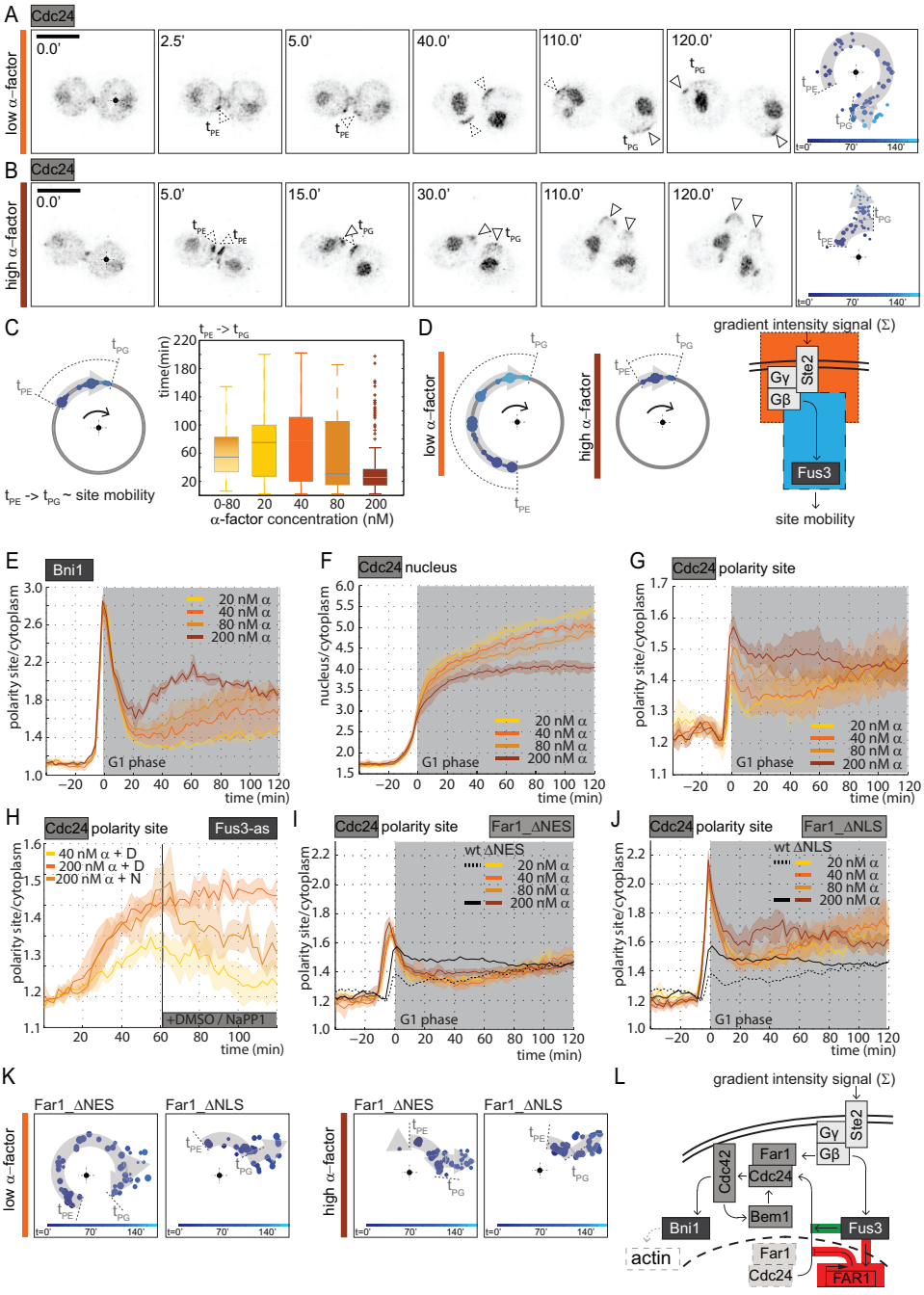
A
Cdc24
low α–factor
0.0'
2.5'
5.0'
40.0'
110.0'
120.0'
B
Cdc24
high α–factor
0.0'
5.0'
15.0'
30.0'
110.0'
120.0'
C
tPE -> tPG ~ site mobility
tPE -> tPG
time(min)
α-factor concentration (nM)
D
low α–factor
high α–factor
gradient intensity signal (Σ)
Gγ
Ste2
Gβ
Fus3
site mobility
E
Bni1
polarity site/cytoplasm
20 nM α
40 nM α
80 nM α
200 nM α
G1 phase
time (min)
F
Cdc24 nucleus
nucleus/cytoplasm
G
Cdc24 polarity site
H
Cdc24 polarity site
Fus3-as
40 nM α + D
200 nM α + D
200 nM α + N
+DMSO / NaPP1
I
Cdc24 polarity site
Far1_ΔNES
wt ΔNES
J
Cdc24 polarity site
Far1_ΔNLS
wt ΔNLS
K
Far1_ΔNES
Far1_ΔNLS
low α–factor
high α–factor
L
gradient intensity signal (Σ)
Gγ
Ste2
Gβ
Cdc42
Far1
Cdc24
Bem1
Bni1
Fus3
actin
Far1
Cdc24
FAR1

Figure 43 *(previous page)*: *Fus3 driven Cdc24 sequestration limits Cdc42 activation and is required for polarity site mobility.* (A & B) Cells expressing Cdc24-qV exposed to 40 nM (low, (A)) or 200 nM (high, (B)) α-factor with the last panel summarizing the time projected position (Fig. 42 D). (C) Extent of polarity site scanning (time from t_{PE} to t_{PG}) in cells expressing Cdc24-qV treated as indicated. (D) The intensity signal controls polarity site mobility. (E) Bni1-qV membrane intensity in cells treated with indicated uniform concentrations of α-factor. Peak at G1 entry corresponds to Bni1 localization to cytokinetic ring. Mean (N = 2, $n_{\emptyset}$ = 130) and standard error of the mean (shaded area) are plotted. (F & G) Cdc24-qV nuclear to cytoplasmic ratio (F) and polarity site intensity (G) in cells treated with indicated uniform concentrations of α-factor (N = 3, $n_{\emptyset}$ = 128). (H) Cdc24-qV nuclear to cytoplasmic ratio in cells expressing a Fus3-analogue sensitive allele (-as) and treated as indicated (D = DMSO, N = 0.1 µM NaPP1, N = 3, $n_{\emptyset}$ = 47). (I & J) Cdc24-qV nuclear to cytoplasmic ratio in cells expressing Far1_ΔNES (I) or Far1_ΔNLS (J). (K) Time projected positions of the Cdc24 polarity site from single cells of experiments in (I) & (J). (L) Scheme summarizing polarity activation control (see text for details).

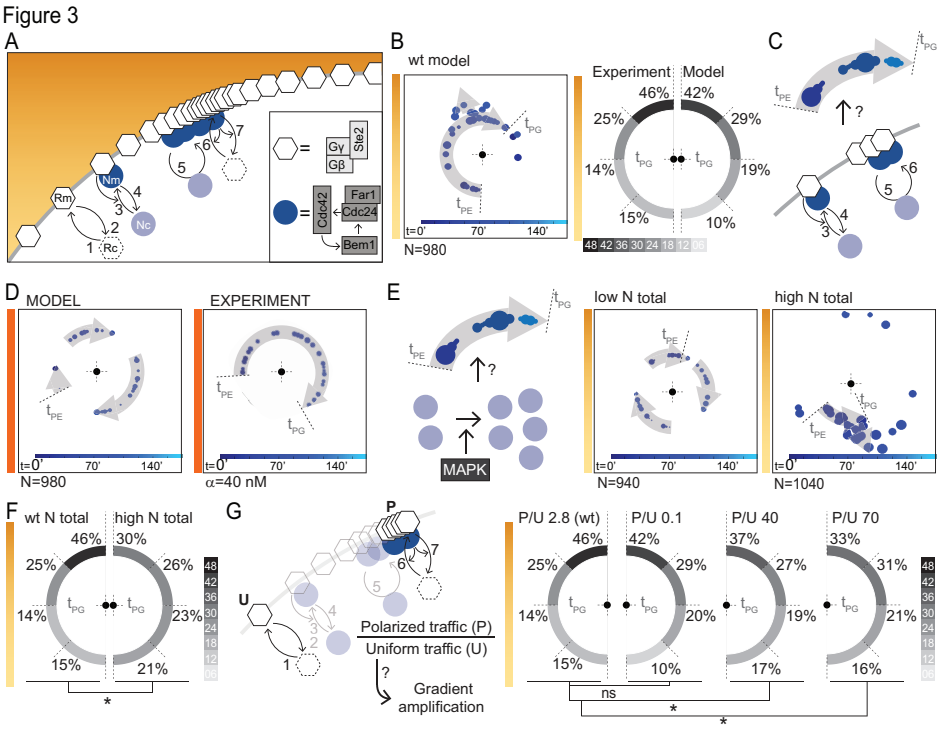

Figure 44: *Limited GTPase activation is sufficient for polarity site mobility and enables a spatial double positive feedback system for gradient amplification.* (A) Computational model incorporating a spatial double positive feedback loop between the polarity complex (N) and the receptor (R). (B) Is double positive feedback sufficient to guide polarity towards the gradient? Polarity site time projected position (Fig. 42 D) from a single model run and direction at t_{PG} of modelled (960 simulations) or experimental (see Fig. 42 F) data. (C & D) Is single component positive feedback sufficient for a fluctuating polarity site (C). Time projected position of Nm from a single model run in uniform α-factor (D). (E & F) Is control of available N levels required for gradient sensing? Shown are time projected positions from single simulations (E) and direction at t_{PG} for model run with high N (F). (G) Is regulation of the ratio of polarized (P) to unpolarized (U) receptor traffic important for gradient amplification? Direction at t_{PG} for ratios below (0.1) or above (40, 70) wt conditions are shown. Two-sample Kolmogorov-Smirnov test vs. wt showed significant difference for ratios 40 ($p=0.016$) and 70 ($p=0.007$). All data are from 960 simulations.

Figure 4

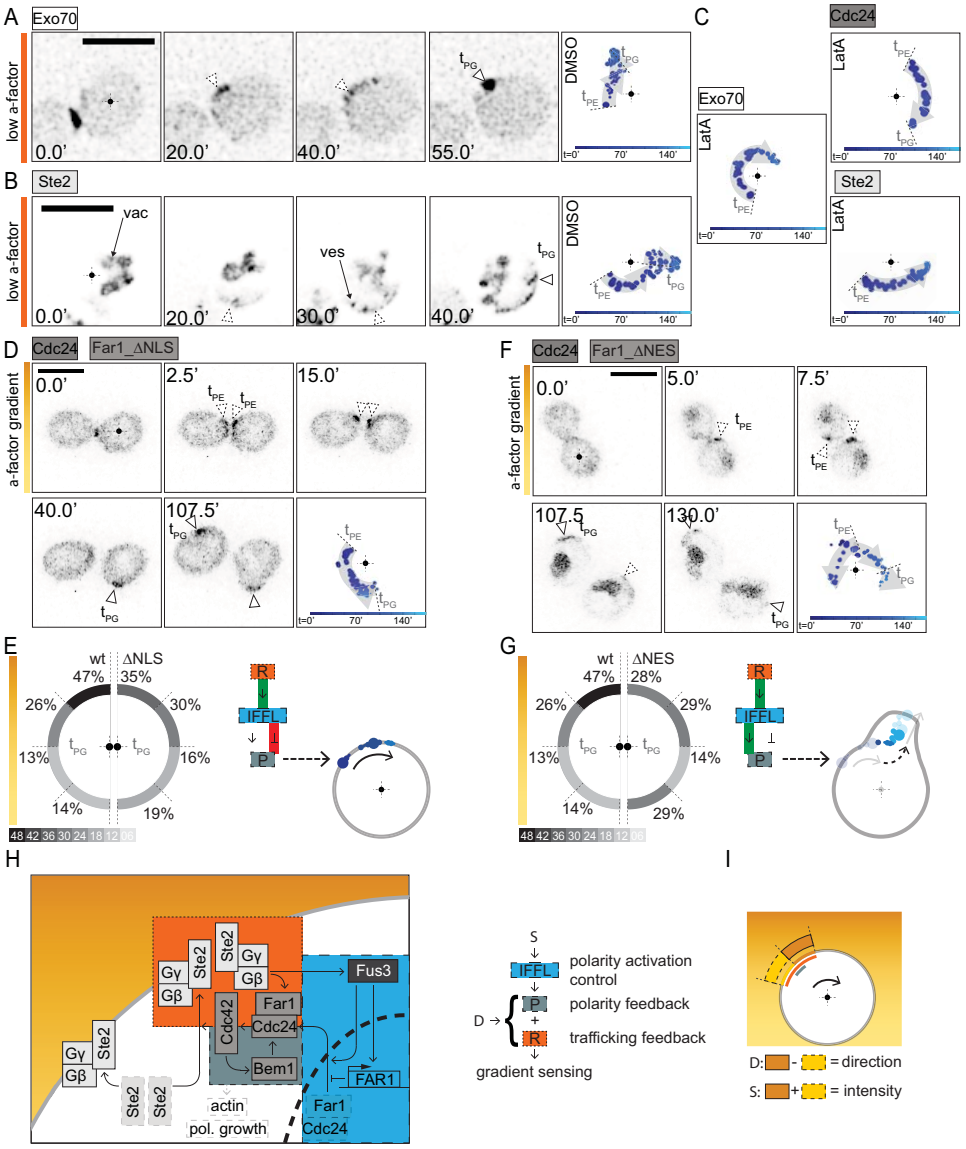
A
Exo70
low a-factor
0.0'
20.0'
40.0'
55.0'
DMSO
B
Ste2
vac
ves
C
Cdc24
LatA
D
Cdc24
Far1_ΔNLS
a-factor gradient
2.5'
15.0'
107.5'
F
Far1_ΔNES
5.0'
7.5'
130.0'
E
wt
ΔNLS
47%
35%
26%
30%
13%
16%
14%
19%
IFFL
G
ΔNES
28%
29%
14%
H
Fus3
Far1
Cdc24
Cdc42
Bem1
FAR1
actin
pol. growth
I
polarity activation control
polarity feedback
trafficking feedback
gradient sensing
D: - = direction
S: + = intensity

Figure 45 *(previous page)*: *Actin independent receptor trafficking and Fus3-activated polarity are key quantitative links building the gradient decoding network.* (A & B) Stills and time projected polarity site positions (Fig. 42 D) of cells expressing an exocytosis marker (Exo70-GFP (A))or Ste2-qV (B) exposed to 40 nM α-factor and DMSO. (C) Time projected position of the polarity site for cells expressing Cdc24-qV, Exo70-GFP or Ste2-qV exposed to 40 nM α-factor and LatA (D) Stills and time projected polarity site positions of cells expressing Cdc24-qV and Far1_ΔNLS exposed to microfluidic α-factor gradient from 0-80 nM. (E) Population mean of direction at t_{PG} for cells treated as in (D) (wt: N = 15, $n_{\emptyset}$ = 40; ΔNLS: N = 6, $n_{\emptyset}$ = 112). (F) Stills and time projected polarity site positions of cells expressing Cdc24-qV and Far1_ΔNES exposed to microfluidic α-factor gradient from 0-80 nM. (G) Population mean of direction at t_{PG} for cells treated as in (F) (wt: N = 15, $n_{\emptyset}$ = 40; ΔNES: N = 2, $n_{\emptyset}$ = 70). (H) Scheme depicting the identified gradient decoding signaling network as explained in the text. (I) Localized *front-back* difference calculation in successive decision steps as explained in the text.

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY FIGURES

Figure S1

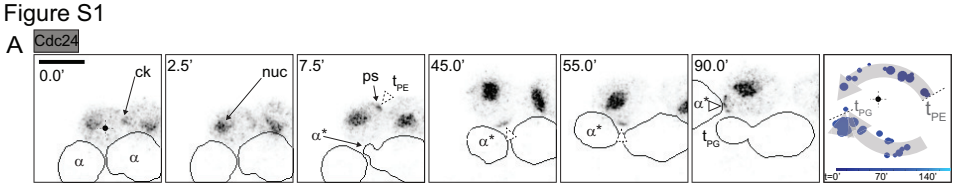

Figure 46: (A) a-type cells expressing Cdc24-qV mixed with wt α-type cells (α, outline drawn from bright field image). Indicated are site of cytokinesis (ck), nuclear Cdc24 localization (nuc), Cdc24 polarity site (p). Dashed arrow heads show polarity establishment (t_{PE}) and scanning site, closed arrow heads indicate onset (t_{PG}) and cell-cell fusion. $\alpha^\star$ marks the future partner cell. Last panel represents the projected position (circle relative to bull's eye (ref)), intensity (circle size) and time (color) of the Cdc24 polarity site.

Figure S2

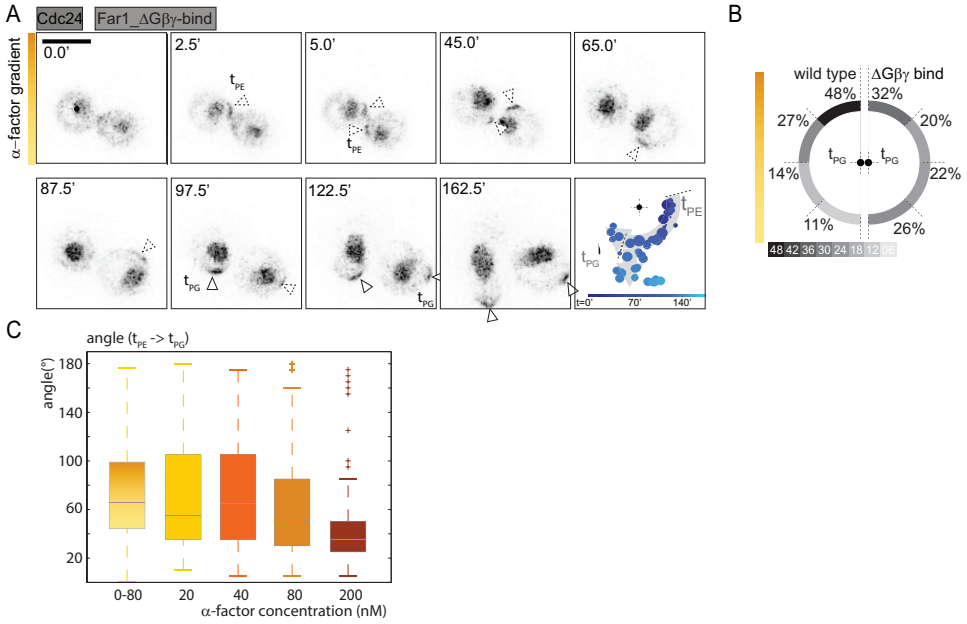

Figure 47: (A) Cells expressing Cdc24-qV and a mutant version of Far1 not able to bind the Gβγ heterodimer of the activated receptor (Far1_ΔGβγ-bind) were exposed to an α-factor gradient of 0 to 80 nM. Time projected positions of the Cdc24 polarity site is shown in last frame. (B) Angle of the polarity site with respect to the gradient was determined at t_{PG} for cells treated as described in (A), results binned in 45° increments and expressed as a percentage (N = 2, $n_{\emptyset}$ = 102). Mutant cells were mixed with unlabeled wt cells in the same chamber to control for gradient stability. (C) Angle (irrespective of gradient direction) between polarity site position at establishment of polarity (t_{PE}) and polarity site position at initiation of polarized growth (t_{PG}) quantified from cells exposed to a 0-80 nM α-factor gradient (same experiment as described in Fig. 42 D) or to uniform concentrations of α-factor (experiment as described in Fig. 43 A and B). Note that in low uniform α-factor concentrations, although there is no gradient present, polarized growth is initiated at a similar variable distance away from the site of polarity establishment as in cells exposed to pheromone gradients.

Figure S3

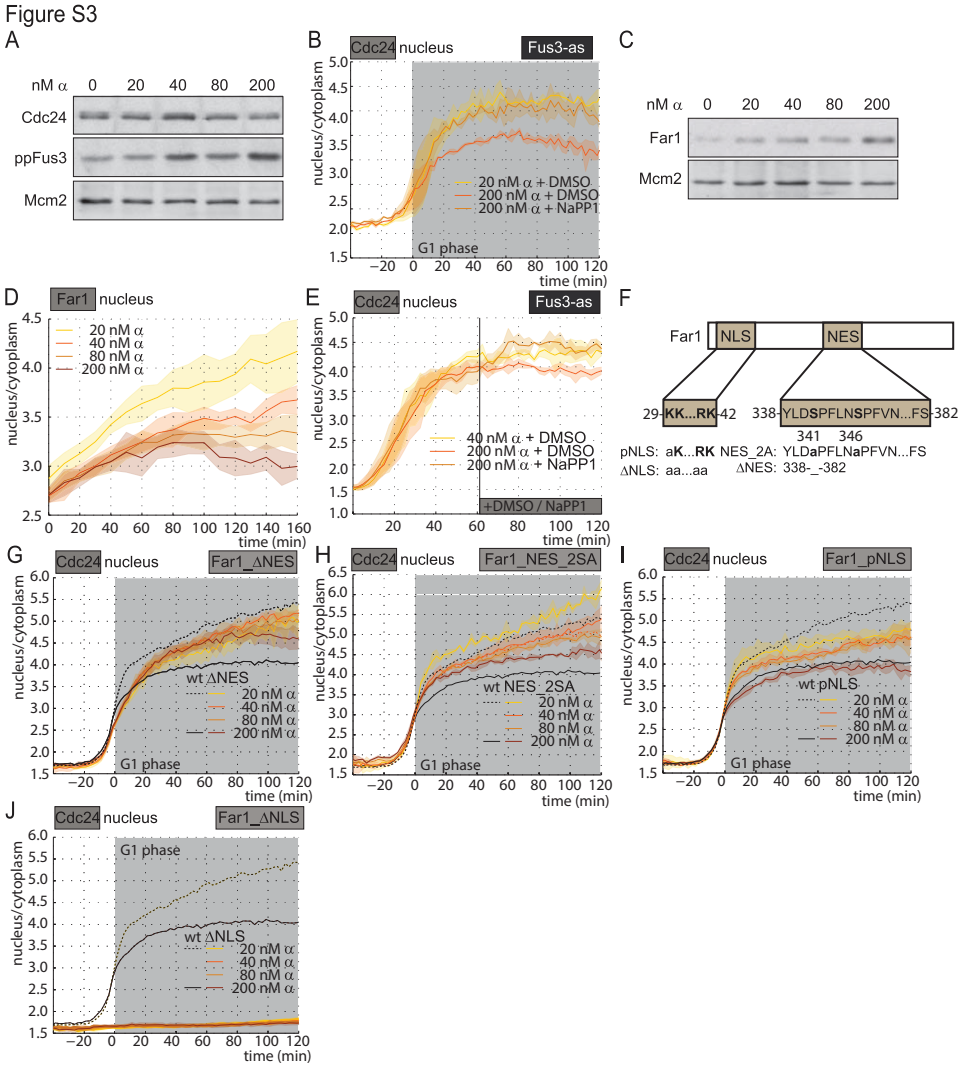
A
nM α 0 20 40 80 200
Cdc24
ppFus3
Mcm2
B
Cdc24 nucleus
Fus3-as
nucleus/cytoplasm
20 nM α + DMSO
200 nM α + DMSO
200 nM α + NaPP1
G1 phase
time (min)
C
nM α 0 20 40 80 200
Far1
Mcm2
D
Far1 nucleus
nucleus/cytoplasm
20 nM α
40 nM α
80 nM α
200 nM α
time (min)
E
Cdc24 nucleus
Fus3-as
nucleus/cytoplasm
40 nM α + DMSO
200 nM α + DMSO
200 nM α + NaPP1
+DMSO / NaPP1
time (min)
F
Far1
NLS
NES
29-KK...RK-42
338-YLDSPFLNSPFVN...FS-382
341 346
pNLS: aK...RK NES_2A: YLDaPFLNaPFVN...FS
ΔNLS: aa...aa ΔNES: 338-_-382
G
Cdc24 nucleus
Far1_ΔNES
nucleus/cytoplasm
wt ΔNES
20 nM α
40 nM α
80 nM α
200 nM α
G1 phase
time (min)
H
Cdc24 nucleus
Far1_NES_2SA
nucleus/cytoplasm
wt NES_2SA
20 nM α
40 nM α
80 nM α
200 nM α
G1 phase
time (min)
I
Cdc24 nucleus
Far1_pNLS
nucleus/cytoplasm
wt pNLS
20 nM α
40 nM α
80 nM α
200 nM α
G1 phase
time (min)
J
Cdc24 nucleus
Far1_ΔNLS
nucleus/cytoplasm
G1 phase
wt ΔNLS
20 nM α
40 nM α
80 nM α
200 nM α
time (min)

Figure 48 *(previous page)*: (A) Cells expressing Cdc24-GFP were S-phase synchronized, released into fresh medium containing indicated concentrations of a-factor for 100 minutes, trichlor acid (TCA) fixed and whole cell extract prepared. Indicated proteins were detected by western blotting. Representative experiment of three replicates. (B) Quantification of nuclear Cdc24 levels in cells expressing Cdc24-qV and Fus3-as treated with indicated concentrations of α-factor and DMSO or 0.1 µM NaPP1 to partially inhibit Fus3-as (N = 2, $n_{\emptyset}$ = 53). (C) Cells expressing Far1-GFP were S-phase synchronized, released into fresh medium containing indicated concentrations of a-factor for 100 minutes, TCA fixed and whole cell extract prepared. Indicated proteins were detected by western blotting. Representative experiment of three replicates. (D) Quantification of nuclear Far1 levels in cells expressing Far1-qV treated with indicated uniform α-factor concentrations (N = 4, $n_{\emptyset}$ = 57). (E) Quantification of nuclear Cdc24 levels in cells expressing Cdc24-qV and Fus3-as treated with indicated concentrations of α-factor and a pulse of DMSO or 0.1 µM NaPP1 to partially inhibit Fus3-as (N = 3, $n_{\emptyset}$ = 47). (F) Far1 bi-partite Nuclear Localization Signal (NLS) and Nuclear Export Signal (NES) with mutated wt residues indicated in bold and mutant sequence in small letters. _ represents deletion. (G) Nuclear to cytoplasmic ratio of cells expressing Cdc24-qV and Far1_ΔNES treated with indicated uniform α-factor (N = 4, $n_{\emptyset}$ = 80). (H) Quantification of nuclear to cytoplasmic levels for cells expressing Cdc24-qV and Far1-NES2SA exposed to indicated concentrations of uniform α-factor (N=2, $n_{\emptyset}$ = 87). (I) Quantification of nuclear to cytoplasmic levels for cells expressing Cdc24-qV and Far1-pNLS exposed to indicated concentrations of uniform α-factor (N=2, $n_{\emptyset}$ = 89). (J) Quantification of nuclear Cdc24 levels in cells expressing Cdc24-qV and Far1_ΔNLS treated with indicated concentrations (N = 4, $n_{\emptyset}$ = 73).

Figure S4

A

Rct	Description	Rate (min^{-1} x mol)	Rate (N=980) (min^{-1} x Xm)
1	Rc -> Rm	0.01735	0.01735
2	Rc <- Rm	1.735	1.735
3	Rm + Nc -> Rm + Nm	0.00005	0.005
4	Nm <- Nc	9	9
5	Nc + Nm -> Nm + Nm	0.01	0.7
6	Rc + Nm -> Rm + Nm	0.000694	0.04858
7	Rc + Nm <- Rm + Nm	0.0694	4.858

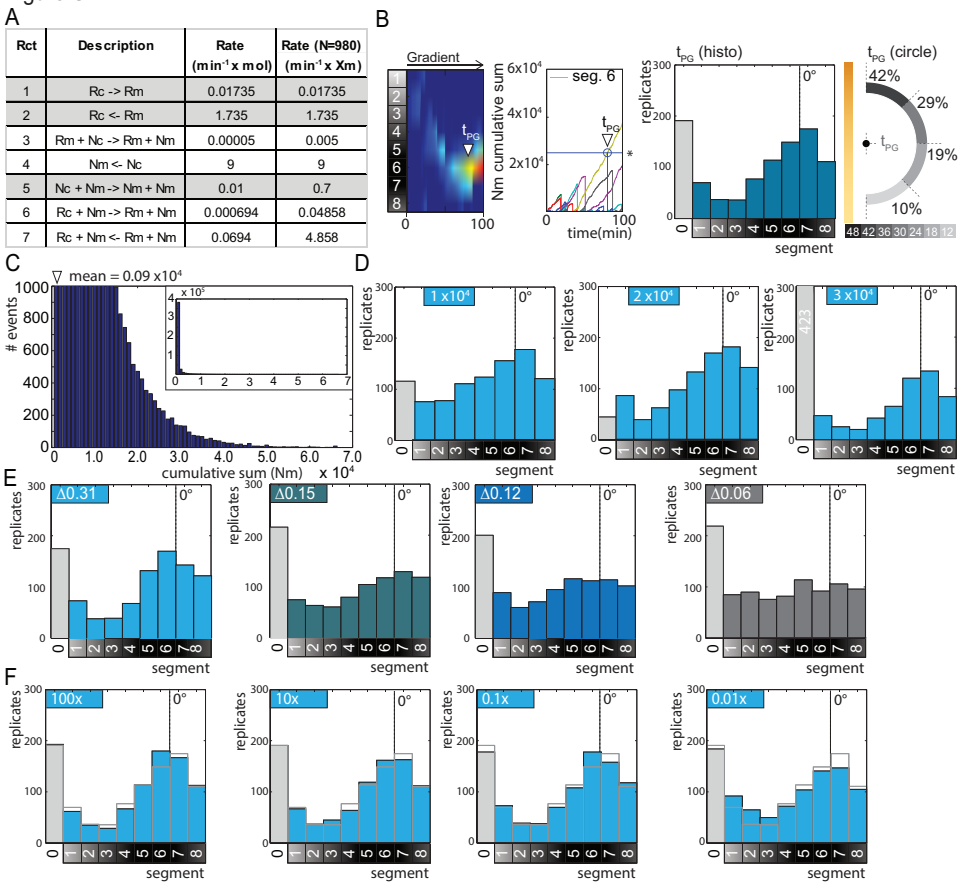

Figure 49 *(previous page)*: (A) Reactions, rate constants and rates used for gradient decoding model. Rates for N=980 with steady state levels of Nm (70) and Rm (100) affecting the frequency of reactions 3, 5, 6 and 7 are shown. (B) For each simulation, we determine in which segment the polarity site stabilizes using a threshold ($\star$) of accumulated Nm per segment (see Methods for details). Determined t_{PG} is indicated. Direction of t_{PG} is plotted from all 960 simulation as histogram (histo) or as circle (circle) representation. Segment 0 counts simulations where the Nm threshold was not reached within 300 minutes. Gradient direction (0°) is indicated by dashed line (histo) or shaded bar (circle). (C) Distribution of Nm cumulative sum for 960 replicates of a model run at N=980. The mean Nm cumulative sum is indicated and the inset shows the same histogram at a full y-axis scale up to 4x 105. D) Histogram of segments reaching the Nm threshold across 960 replicates using three different thresholds for the Nm cumulative sum. At 3x104 the threshold is not reached in 423 replicates. Gradient direction (0°) is indicated by dashed line. E) Segment histograms for simulations using different gradient steepness (fractional difference of activated receptors between front and back is indicated). F) Segment histograms for simulations using a higher or lower overall rate for all exocytosis and endocytosis reactions. Note how gradient sensing is stable across a large range of trafficking rates and only slightly reduced at very low trafficking rates. Rm steady state in all cases is unchanged (Rm = 99, Nm = 72).

Figure S5

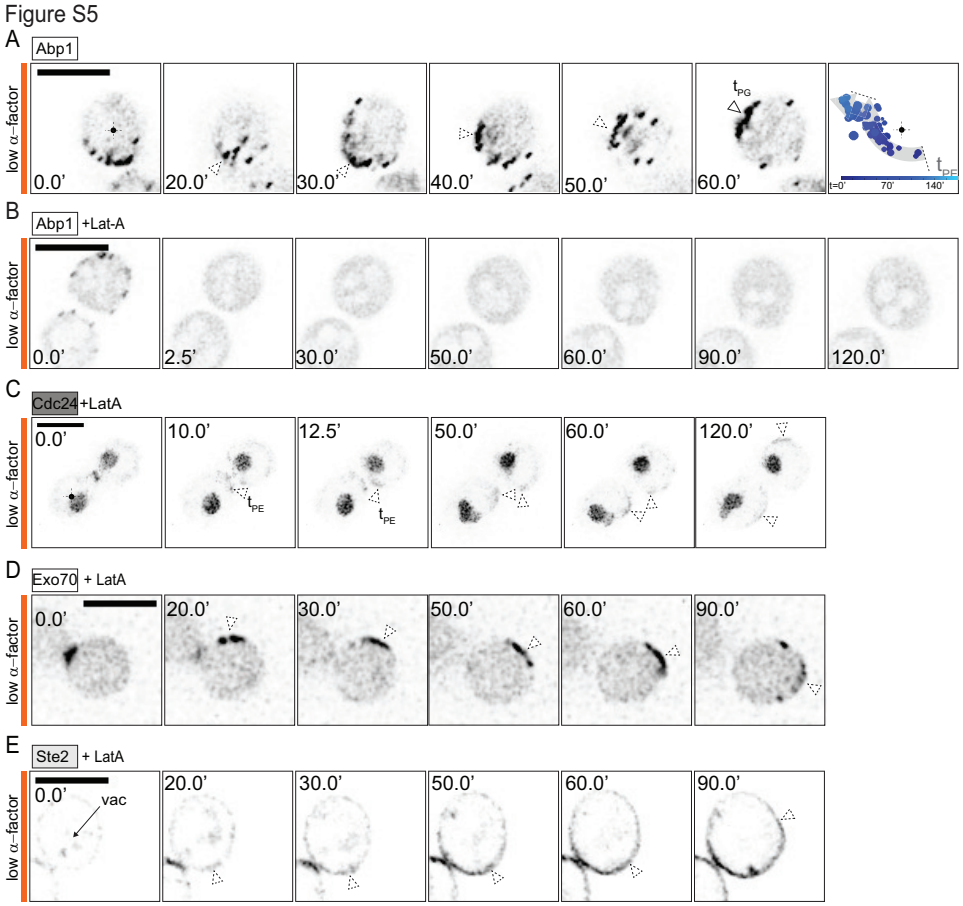

Figure 50: (A & B) Cells expressing the endocytic marker Abp1-GFP were exposed to 40 nM uniform α-factor and DMSO (A) or 200 µM LatA (B) just before finishing cytokinesis. Time projected positions of the Abp1 foci center (marked by arrow in images). Note how LatA treatment quickly disperses all Abp1 foci, confirming complete actin depolymerization. (C) Cells expressing Cdc24-qV were exposed to 40 nM uniform α-factor and 200 µM LatA just before finishing cytokinesis. (D) Cells expressing the exocytosis marker Exo70-GFP were exposed to 40 nM uniform α-factor and 200 µM LatA just before finishing cytokinesis. (E) Cells expressing Ste2-qV were exposed to 40 nM uniform α-factor and 200 µM LatA just before finishing cytokinesis. vac = vacuolar Ste2-qV.

SUPPLEMENTARY MATERIALS

MATERIALS AND METHODS

YEAST STRAINS, PLASMIDS AND GROWTH CONDITIONS

See Table 14 for a list of Plasmids. Point mutants in Far1 were generated using the QuickChange system (Agilent Technologies) on plasmids designed to replace FAR1 at its endogenous location (see Table 14 for a list of Plasmids). The Far1-NES (amino acids 338-382) was deleted using synthetic DNA technology and a plasmid replacing GFP fusions from the genome-wide library [104] with qV was constructed based on a previously published replacement plasmid [235].

All yeast strains are derivatives of BY4741 [30], all fusion proteins are expressed from their endogenous location and strains are listed in Table 15. Gene fusions were generated by homologous recombination-based replacement of the endogenous gene.

Strains for all experiments were grown in SC-based media (0.17% Yeast nitrogen base, 2% glucose, 0.5% NH_4-sulfate and amino acids). Cell synchronization was performed by adding 100 µM Hydroxyurea (HU) to logarithmically growing strains for 2 hours. HU was removed by filtration and extensive washing in pre-warmed media. Filtered cells were placed in fresh, pre-warmed media and typically exited mitosis within 60 to 90 minutes post release. α-factor was dissolved in solutions conaining 2% BSA and 50% glycerol, and added to growth media supplemented with 134 µg ml^{-1} BSA to avoid unspecific binding of α-factor to surfaces.

PROTEIN EXTRACTS AND WESTERN BLOTTING

Protein extracts were prepared from Trichlor Acid (TCA) fixed cells. TCA pellets were resuspended in 2x Urea buffer (62.5 mM Tris pH 6.8, 10% glycerol, 4% SDS, 5% β-mercaptoethanol, 8 M Urea and bromophenol blue) and vortexed 3 minutes at 4° C with 0.5 mm glass beads. After boiling, samples were analyzed using standard SDS-PAGE and western blotting procedures. Antibodies used were: α-GFP (11 814 460 001, Roche), α-Phospho-p44/42 Antibody (#9101, Cell Signaling), α-Mcm2 (sc-6680, Santa Cruz).

MICROSCOPY AND MICROFLUIDICS

Images were acquired on fully automated inverted epi-fluorescence microscopes (Ti-Eclipse, Nikon) in an incubation chamber set to 30° C, with 60x oil objectives and appropriate excitation and emission filters. A motorized XY-stage and piezo drive was used to acquire z-stacks (8 steps at 0.8 µM) and multiple fields of view per time point. Cells for gradient experiments were imaged in homemade microfluidic gradient chips and experimental

cells were always mixed with unlabeled wt cells to control for gradient stability, for dose response experiments in Cellasics Y04C chips (Millipore Corp.) and for Fus3-as experiments in 96-well 17 mm glass-bottom plates (Matrical Corp.). Except for Cellasics chips, slides were pre-coated using a filtered solution of Concanavalin A in PBS (1 mg ml^{-1}).

IMAGE ANALYSIS

Automated image analysis was performed using YeastQuant software on raw images [175] (37) running in Matlab. Cells were segmented using double-Cherry fused to the transmembrane domain of SNC2 (TMD-Cherry, marking cortical ER) and split into the separate objects of perimeter (= TMD-Cherry signal and 3 pixels outside of it), rim (= cytoplasmic regions within 4 pixels inside the membrane) and nucleus (region more than 5 pixels inside the membrane, i.e. mostly the cell nucleus). Nuclear to cytoplasmic ratio corresponds to the 20 highest pixels (HiPix) of the nucleus object divided by the average intensity of the rim object. Polarity site intensity corresponds to HiPix of perimeter divided by the average intensity of the rim object. Positional information was obtained using Pie Segmentation to cut each object into 72 slices. Morphological information was extracted by calculating the distance from cell border to cell center (as defined by fitting a circle into the already segmented cell) for each of the 72 slices. Initiation of polarized growth was identified by an increase in the ratio of maximal to minimal distance across all 72 slices. Images for display were deconvolved using the Huygens software (SVI) and projected using maximum intensity projection unless otherwise indicated. All data plots were generated using the Matlab statistics toolbox. Cell and polarity site orientation for experiments in α-factor-gradients was quantified manually using Fiji [197].

COMPUTATIONAL MODELING

Our stochastic model is based on the implementation of a polarization model [112] and aims at qualitative simulation of gradient sensing between t_{PE} and t_{PG}. It consists of four species (cytoplasmic polarity complex Nc, membrane bound active polarity complex Nm, cytoplasmic receptor complex Rc, membrane bound active receptor complex Rm) and seven reactions as given in Fig. 49 A. Each simulation was run with 960 replicates. Total number of N and R is constant throughout each simulation with Ntotal as indicated. We estimated the levels of R based on the following consideration. Upon α-factor stimulation ligand bound Ste2 is immediately endocytosed [196] and degraded [98], newly synthesized Ste2 molecules are transported to the membrane by exocytosis [111], resulting in extensive changes of receptor levels in the cell membrane. Since our model considers the steady state after signaling is initiated, we set the steady state level of R to 10000.

Membrane bound molecules have position values in the continuous interval [0,1] assigned, and the cytoplasm is considered well mixed, thus reactions are based on the law of mass action. Spontaneous recruited receptors are assigned a position dependent on a gradient of α-factor. Due to the high affinity of α-factor binding to the receptor ($\approx$ 6 nM [110]), receptor exocytosis and activation were combined into one reaction. Thus, all membrane bound receptors are active.

Rm recruits Nc (Far1-mediated recruitment [34, 163]) and Nm recruits Rc (polarized exocytosis). New membrane bound complex is assigned the position of the recruiting molecule. Membrane diffusion is calculated for each molecule individually after each reaction in the simulation, with d_Rm = 1/10 d_Nm [112, 234]. Nm dissociates Rm in proximity of Nm (polarized endocytosis). Molecule abundance was simulated using an implementation of Gillespie's stochastic simulation algorithm.

Since our model is not designed to simulate polarized growth, we needed to establish a heuristic approach to identify when polarized growth is initiated. Manual inspection of simulation runs revealed that polarity sites displayed a longer life time and less lateral movement towards the gradient than away from the gradient. To quantify this polarity site stabilization in a model run, we divide the cell membrane into 8 equally sized segments. For each segment, we calculate the cumulative sum of polarity complexes Nm over time. The cumulative sum for each segment is reset when no Nm, and thus no polarity site, is present in the respective segment. A stable polarity site should reach a set threshold of the Nm cumulative sum. We tested different Nm cumulative sum thresholds based on multiples of the mean of all integration events (0.09 x 104, Fig. 49 C). Within a range of 2 x 104 to 3 x 104, the segment distribution reflected our manual analysis with more stable polarity sites on the up-gradient segments (Fig. 49 D). Below this range the included polarity site were equally distributed, inconsistent with our manual analysis; while above this range a large fraction of polarity sites were excluded from the analysis (> 40%) introducing a strong bias. We thus used a threshold of 2.5 x 10^4 Nm cumulative sum for defining the time when polarized growth is initiated.

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY TABLES

PLASMID ID	GENOTYPE	SOURCE
pBH94	CDC24-qV-URA3	this study
pBH80	FAR1-qV-URA3	this study
pBH98	FAR1-2HA-Strp-HIS3	this study
pBH132	FAR1_C205Y-2HA-Strp-HIS3 = ΔGβ-bind	this study
pBH246	FAR1-NLS3A-2HA-Strp-HIS3 = pNLS	this study
pBH229	FAR1-NLS4A-2HA-Strp-HIS3 = ΔNLS	this study
pBH230	FAR1-ΔNES-2HA-Strp-HIS3 = ΔNES	this study
pBH248	FAR1-S341A_S346A-2HA-Strp-HIS3 = NES2SA	this study
pBH118	GFPhom-2HA-qV-URA3	this study
pSP160	pRPS2_dCherry-TMD	S. Pelet

Table 14: Plasmids

STRAIN ID	GENOTYPE	SOURCE
yBH66	BY4741 *(his3Δ1; LEU22Δ0; met15Δ0; URA33Δ0; MATa)*	OpenBiosystems
yBH67	BY4742 *(his3Δ1; LEU22Δ0; lys2Δ0; URA33Δ0; MATα)*	OpenBiosystems
yBH402	*BNI1-qV::URA; TMD-dCherry::LEU2*	this study
yBH197	*CDC24-qV::HIS; fus3-as::URA3; TMD-dCherry::LEU2*	this study
yBH217	*FAR1-qV::URA3; TMD-dCherry::LEU2*	this study
yBH203	*STE2-qV::URA; TMD-dCherry::LEU2*	this study
yBH405	*TMD-dCherry::LEU2*	this study
yBH469	*CDC24-qV::URA; TMD-dCherry::LEU2; FAR1-HAS::HIS*	this study
yBH473	*CDC24-qV::URA; TMD-dCherry::LEU2; far1-c205y-HAS::HIS*	this study
yBH533	*CDC24-qV::URA; TMD-dCherry::LEU2; far1-nls3A-HAS::HIS*	this study
yBH508	*CDC24-qV::URA; TMD-dCherry::LEU2; far1-nls4A-HAS::HIS*	this study
yBH539	*CDC24-qV::URA; TMD-dCherry::LEU2; far1-S341A-S346A-HAS::HIS*	this study
yBH529	*CDC24-qV::URA; TMD-dCherry::LEU2; far1-Δnes-HAS::HIS*	this study
yBH196	*ABP1-GFP::HIS*	OpenBiosystems
yBH544	*EXO70-GFP::HIS*	OpenBiosystems

Table 15: Yeast Strains

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY MOVIE CAPTIONS

All movies (except where indicated) are deconvolved, projected and inverted z-stacks acquired as described in the methods section, the scale bar is 5 µm. Cytokinesis is always set as t=0′. For feature annotation please see the indicated main figure.

Movie S1

Full movie for stills displayed in Figure 1D. a-type cells expressing Cdc24 qV exposed to α-factor gradient in a microfluidic gradient chip from 0 to 80 nM (Δ 2.5 nM across the cell).

Movie S2

Full movie for stills displayed in Figure S1A. a-type cells expressing Cdc24 qV mixed with α cells and followed to cell fusion (t=92.5).

Movie S3

Full movie for stills displayed in Figure S2A. Cells expressing Cdc24-qV and a mutant version of Far1 not able to bind the Gβ subunit of the activated receptor (ΔGβ-bind) were exposed to an α-factor gradient of 0 to 80 nM.

Movie S4

Full movie for stills displayed in Figure 2A. Cells expressing Cdc24-qV exposed to 40 nM uniform alpha factor.

Movie S5

Full movie for stills displayed in Figure 4A. Cells expressing Exo70-GFP were treated with 40 nM uniform α-factor. Images were acquired in single focal plane, deconvolved and inverted.

Movie S6

Full movie for stills displayed in Figure 4B. Cells expressing Ste2-qV were treated with 40 nM uniform α-factor. Images were acquired in single focal plane, deconvolved and inverted.

Movie S7

Full movie for stills displayed in Figure S5A. Cells expressing Abp1-GFP were treated with 40 nM uniform α-factor. Images were acquired in single focal plane, deconvolved and inverted.

Movie S8

Full movie for stills displayed in Figure S5C. Cells expressing Cdc24-qV were treated with 40 nM uniform α-factor and 200 µM LatA just before finishing cytokinesis.

Movie S9

Full movie for stills displayed in Figure S5D. Cells expressing Exo70-GFP were treated with 40 nM uniform α-factor and 200 µM LatA just before finishing cytokinesis. Images were acquired in single focal plane, deconvolved and inverted.

Movie S10

Full movie for stills displayed in Figure S5E. Cells expressing Ste2-qV were treated with 40 nM uniform α-factor and 200 µM LatA just before finishing cytokinesis. Images were acquired in single focal plane, deconvolved and inverted.

Movie S11

Full movie for stills displayed in Figure 4D. Cells expressing Cdc24-qV and Far1-ΔNLS were subjected to a 0 to 80 nM α-factor gradient and followed from cytokinesis to after polarized growth was initiated.

Movie S12

Full movie for stills displayed in Figure 4F. Cells expressing Cdc24-qV and Far1-ΔNES were subjected to a 0 to 80 nM α-factor gradient and followed from cytokinesis to after polarized growth was initiated.

Part IV

WISDOM OF THE CROWD: SCIENTIFIC CHALLENGES FOR THE EVALUATION OF METHODS IN SYSTEMS BIOLOGY

9

MANUSCRIPT 7

REFERENCE

P. Nandy, M. Unger, C. Zechner, K. K. Dey, and H. Koeppl, *Learning diagnostic signatures from microarray data using L1-regularized logistic regression*, Systems Biomedicine, vol. 1, no. 4, p. e25271, 2013.

AUTHOR CONTRIBUTIONS

PN, MU, CZ and KD implemented the algorithm. PN, MU, CZ and HK conceived the approach and classification pipeline and wrote the paper.

LEARNING DIAGNOSTIC SIGNATURES FROM MICROARRAY DATA USING L1-REGULARIZED LOGISTIC REGRESSION

Preetam Nandy[1], Michael Unger[1], Christoph Zechner[1], Kushal K Dey[2], and Heinz Koeppl[1, 3]

ABSTRACT

Making reliable diagnoses and predictions based on high-throughput transcriptional data has attracted immense attention in the past few years. While experimental gene profiling techniques – such as microarray platforms – are advancing rapidly, there is an increasing demand of computational methods being able to efficiently handle such data. In this work we propose a computational workflow for extracting diagnostic gene signatures from high-throughput transcriptional profiling data. In particular, our research was performed within the scope of the first sbv IMPROVER challenge. The goal of that challenge was to extract and verify diagnostic signatures based on microarray gene expression data in four different disease areas: Psoriasis, Multiple Sclerosis, Chronic Obstructive Pulmonary Disease and Lung Cancer. Each of the different disease areas is handled using the same three-stage algorithm. First, the data is normalized based on a multi-array average RMA normalization procedure to account for variability among different samples and datasets. Due to the vast dimensionality of the profiling data, we subsequently perform a feature pre-selection using a Wilcoxon's rank sum statistic. The remaining features are then used to train an L1-regularized logistic regression model which acts as our primary classifier. Using the four different datasets, we analyze the proposed method and demonstrate its use in extracting diagnostic signatures from microarray gene expression data.

KEYWORDS

classification, gene expression, L1-regularization, LASSO, logistic regression, microarray data, RMA normalization, Wilcoxon rank sum test

ABBREVIATIONS

Adenocarcinoma (AC); Area Under Precision-Recall curve (AUPR); AUPR_{Avg}, average of the AUPR across the classes; Belief Confusion Metric (BCM); Correct Class Enrichment Metric (CCEM); Chronic Obstructive Pulmonary Disease (COPD); Systems Biology Verification combined with Industrial Methodology for Process Verification (sbv IMPROVER); Least Absolute Shrink-

1 BISON Group, Automatic Control Laboratory, ETH Zurich, Zurich, Switzerland
2 Indian Statistical Institute, Kokata, India
3 Correspondence to: Heinz Koeppl; Email: koepplh@ethz.ch

age and Selection Operator (LASSO); Lung Cancer (LC); MS Diagnostic (MSD); Robust Multi-array Average (RMA); Squamous Cell Carcinoma (SCC)

INTRODUCTION

The effective treatment of diseases often relies on making early and accurate diagnoses. However, this can be highly challenging, especially for diseases with complex genetic causes. Microarray techniques are able to capture the expression levels of thousands of genes, opening up a huge source of information about the genetic profiles of patients. While the potential of microarray technologies for medical purposes was repeatedly demonstrated [144, 177, 134], challenges arise in the computational handling of such datasets. Typically, approaches from statistics and machine learning [25, 67] are employed to extract disease-relevant information and to predict diagnostic features such as a patient's disease state. Most of these approaches are *supervised*, meaning that they rely on the availability of labeled training data. [210] Common techniques include linear discriminant analysis, nearest-neighbor classifiers, classification trees, bagging, and boosting [55], support-vector machines [32, 71], neural networks [114], hierarchical Bayesian models [130] and regularized regressions [50, 248].

Typically, the number of case and control samples is just a fraction of the number of probes on a single microarray chip, posing one of the main difficulties in handling such data. Mathematically, the corresponding inverse problems are said to be *ill-posed* or *underdetermined* and their solution requires specialized algorithms. The same situation applies for the data from the first sbv IMPROVER challenge [155, 154], the Diagnostic Signature Challenge. Only a few hundred training samples were provided for each of the four disease datasets, psoriasis, MS Diagnostic (MSD), Chronic Obstructive Pulmonary Disease (COPD) and Lung Cancer (LC), in order to train the classifiers. Based on those, the goal was to predict the disease-probabilities of additional samples from an unlabeled test dataset.

In this work we lay out a computational workflow, which accounts for the complex nature of the high-dimensional microarray datasets. The validity of the approach is benchmarked using four independent datasets within the scope of the sbv IMPROVER challenge. In particular, we show that the method is able to extract disease-relevant gene profiles and demonstrate its potential in making diagnostic predictions.

RESULTS

The Diagnostic Signature Challenge encompassed four independent classification tasks (sub-challenges), each task corresponded to a particular disease and dataset. Three of the four sub-challenges were designated to distinguish between the disease/non-disease (i.e., binary classification) states. The goal of the fourth task, the lung cancer sub-challenge, was to predict four disease states corresponding to two different cancer types (Adeno-

carcinoma (AC) and Squamous Cell Carcinoma (SCC)) and their respective stages (stages I and II). The performance of each classifier was assessed by estimating its prediction success probability.

Computational Workflow

Although the L1-regularized logistic regression provides a natural mechanism for feature selection and prevention of overfitting (see Materials and Methods), it would require massive amounts of computational resources when directly applied to the high-dimensional dataset. Thus, further preprocessing and data reduction had to be performed. More specifically, we followed a workflow that consisted of three main steps. In step one, we normalized the pooled data (comprising both the training and test datasets) for each of the sub-challenges using a standard Robust Multi-array Average (RMA) normalization procedure [106]. In step two, we significantly reduced the dimensionality of the feature space using a nonparametric method based on the Wilcoxon rank sum test statistic [51, 173]. In step three, the remaining features were used to train an L1-regularized logistic regression model. As indicated above, this approach allows to further reduce the number of features used in the final model [25, 67]. The overall predictor for each disease is a monotonic function of the pre-processed and weighted feature intensities corresponding to the diagnostic signatures. Detailed descriptions of the three individual building blocks can be found in Materials and Methods.

SUB-CHALLENGE	PSORIASIS	MSD	COPD	LC (2 CLASSES)	LC (AC STAGE)	LC (SCC STAGE)
# NUMBER GENES SELECTED	15502	9591	2000* (1152)	3260	2000* (3)	2000* (1012)

Table 16: Number of genes selected by the pre-selection algorithm that correspond to each of the sub-challenges.
MS Diagnostic (MSD); Chronic Obstructive Pulmonary Disease (COPD); Lung Cancer (LC). For the psoriasis and MSD sub-challenges, a large number of genes with significant p-value scores were selected.
*For COPD, LC (AC) and LC SCC, because the number of selected genes was low they were replaced by 2000 of the most significant genes in terms of their p-values. The numbers in brackets are the number of variables (the genes) with p-value less than 0.1/(total number of genes) in the Wilcoxon rank sum test.

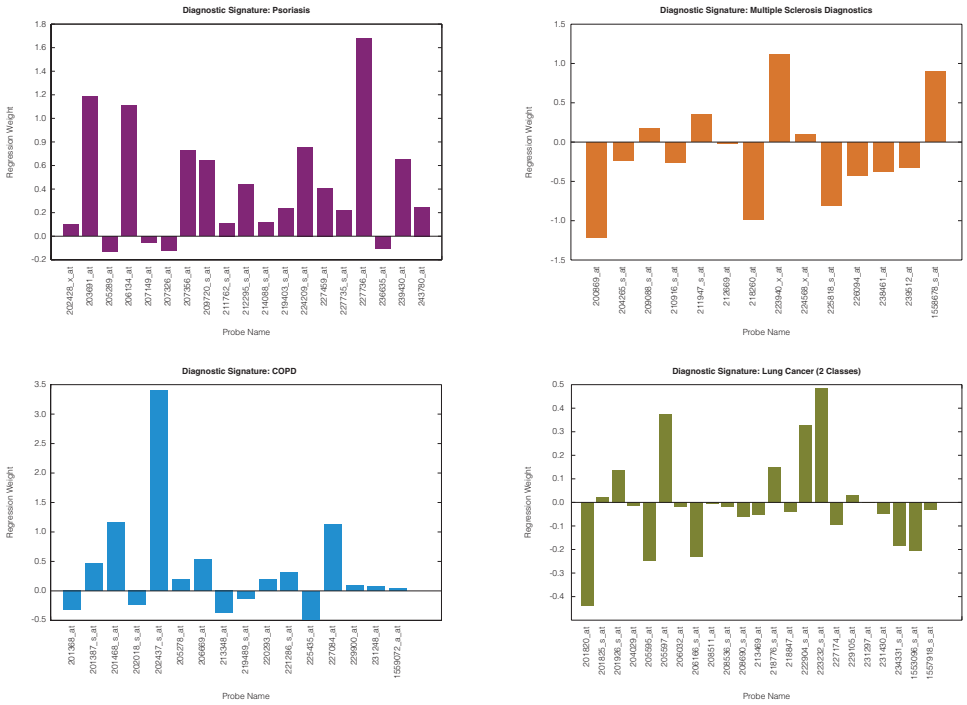

Figure 51: Diagnostic signatures for each of the four sub-challenges. Bar heights indicate how each probe is weighted in the final regressor.

Experimental Results

The numbers of significantly expressed genes revealed by the feature pre-selection algorithm at a 10% level of significance are shown in Table 16. To some extent, the number of pre-selected genes reflects the richness of the disease signature in the expression profiles. Although a large number of pre-selected genes may improve the predictability of the disease state, the complexity of the subsequent classification task increases: the dimensionality becomes large compared to the sample size and standard approaches will inherently suffer from overfitting. Appropriate regularization strategies, such as provided by the L1-regularized logistic regression, can handle such problems to produce more reliable predictions. The selected probe names and their corresponding weights for all four sub-challenges are shown in Figure 51. For each of those, the pre-selection algorithm was able to substantially reduce the number of features and hence, the dimensionality of the resulting dataset. Because all the variables were standardized before training, the absolute weights represent the significance of the corresponding regressor.

Performance measures of our predictions were based on the score values of three sbv IMPROVER standard quality metrics; namely, the Belief Confusion Metric (BCM), the Correct Class Enrichment Metric (CCEM), and the average of the Area Under Precision-Recall curve (AUPR) across the classes (AUPR_{Avg}). Table 17 shows the performance of our predictions according

QUALITY SCORE	(BCM)	(CCEM)	($AUPR_{AVG}$)	RANK OBTAINED
Psoriasis	0.99	0.99	1.00	2
MSD	0.54	0.52	0.62	12*
COPD	0.66	0.68	0.66	4
LC (2 classes)**	0.82	0.84	0.94	N/A
LC (4 classes)	0.43	0.48	0.50	5

Table 17: The quality score values for the three standard quality metrics for each of the sub challenges.
Belief Confusion Metric (BCM); Correct Class Enrichment Metric (CCEM); $AUPR_{Avg}$, average Area Under Precision-Recall curve (AUPR) across the classes; MS Diagnostic (MSD); Chronic Obstructive Pulmonary Disease (COPD); Lung Cancer (LC).
*The original rank was 37. The training dataset that we used for the MSD sub-challenge reported in this paper is different (basically a subset of the one used in the challenge) from that was used in the sbv IMPROVER challenge.
**LC (2classes) was not part of the sbv IMPROVER challenge.

to those score values and the corresponding rank obtained for each of the sub challenges. Psoriasis was predicted well, while the other diseases were not. This might be partially explained by differences in the amount of available training data (see Figure 52). The graphic shows that most training samples where available for the Psoriasis dataset, which ranked best in our study. In contrast, the worst performance was achieved for the MS Diagnostic dataset, associated with a particularly small sample size. However, a variety of other causes might have contributed to the variability in the performance. The tissue used to perform the microarray experiments did not always originate from a location primarily affected by the disease. This might cause strong qualitative differences between the training and test datasets, which might in turn have significant impact on the classification performance.

In order to test for such differences, we evaluated our classifier against the test sets of the respective sub-challenges using a leave-one-out cross-validation[1]. Those results were compared to the original predictions obtained from the training datasets by means of the $AUPR_{Avg}$ metric (see Figure 53). The remaining performance scores are listed in Table 18. In case of the psoriasis data, we observed only minor differences in the performance, even though the classifier was obtained from significantly fewer samples.

In accordance with our hypothesis, a considerable improvement was obtained for the MSD dataset, indicating strong differences between the training and test dataset. Furthermore, when using the latter, the number of available training samples (i.e., $N = 59$) was higher than the original sample size of the training dataset (i.e., $N = 41$).

1 This check was possible only after the gold standard labels of the test samples were published online.

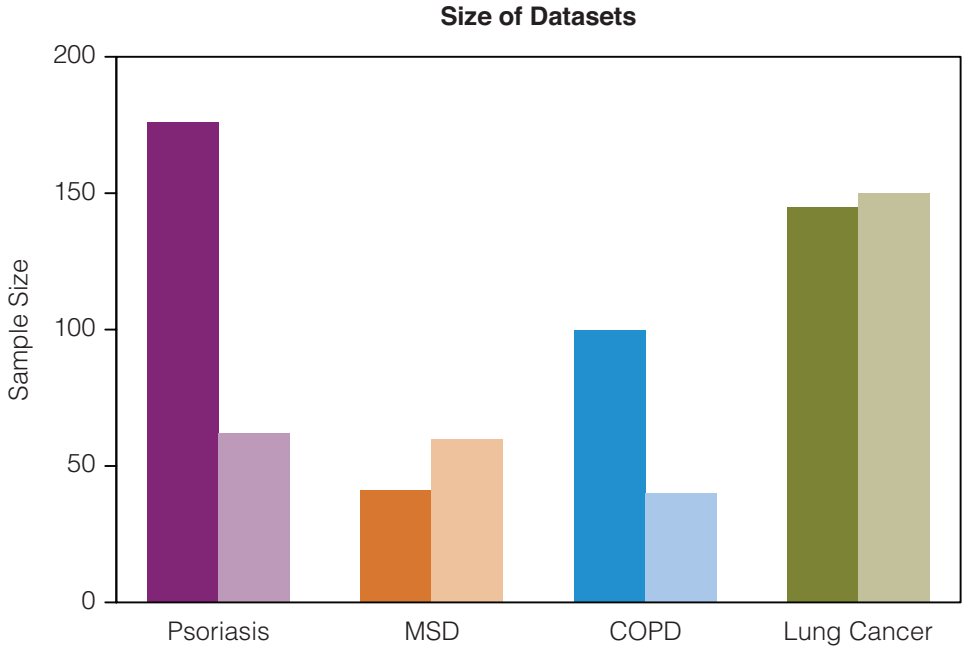

Figure 52: Sizes of the datasets that were available for each of the sub-challenges. Dark bars correspond to training datasets, light bars to test datasets.

For the COPD sub-challenge, the performance of the classifier trained solely on the test dataset was no better than a coin flip (i.e., the success probability was around 0.5). This result suggests that either the available dataset does not contain enough disease-relevant information or the proposed approach is unable to unravel the complexity of the underlying expression patterns. Although COPD is manifested in small airways, the goal was to identify a COPD signature valid for large airways (such as, in this case, the test dataset) for which sample collection is less complex. In case of the training dataset, consisting of samples from both large and small airways, it seems that the classifier was indeed able to extract predictive gene signatures for large airways data.

For the LC sub-challenge, the size of the training set (N = 145) and the size of the test set (N = 150) were roughly the same. However, when LC was considered as a binary classification problem (i.e., classes AC and SCC irrespective of their stage), we found that the classifier performed well in both cases, while for the initial four-class problem (i.e. discriminating between their corresponding stages) the performance was only moderate.

DISCUSSION

In this work we proposed a three-stage computation workflow for extracting diagnostic gene signatures from microarray gene expression data. In order to account for technical and biological variations between individual samples, we first preprocessed the data using a robust multi-array normalization scheme. In order to reduce the dimensionality of the

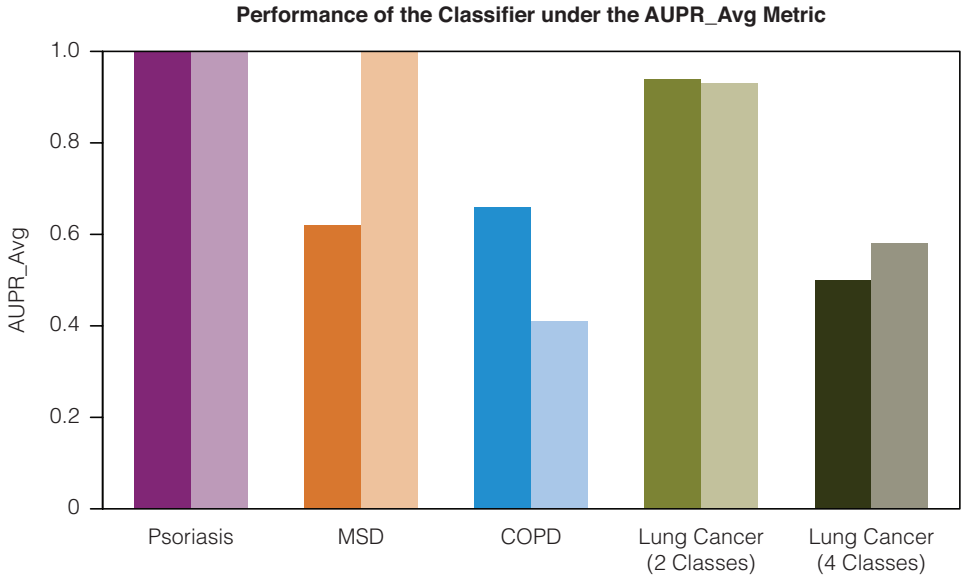

Figure 53: Performance of the Classifier based on the AUPR_{Avg} metric scores. Dark bars correspond to training datasets, light bars to test datasets.

QUALITY SCORE	(BCM)	(CCEM)	(AUPR_{AVG})
Psoriasis	0.99	0.98	1.00
MSD	0.995	0.998	1.00
COPD	0.47	0.48	0.41
LC (2classes)	0.77	0.80	0.93
LC (4classes)	0.48	0.54	0.58

Table 18: The quality score values for the three standard quality metrics for each of the sub challenges.
Belief Confusion Metric (BCM); Correct Class Enrichment Metric (CCEM); AUPR_{Avg}, average Area Under Precision-Recall curve (AUPR) across the classes; MS Diagnostic (MSD); Chronic Obstructive Pulmonary Disease (COPD); Lung Cancer (LC).

datasets, we applied a feature pre-selection algorithm using a Wilcoxon's rank sum statistic. The primary classification algorithm is based on an L1-regularized logistic regression model, which on the one hand is able to prevent overfitting and on the other hand, provides a simple strategy to identify predictive gene signatures. More specifically, the regression weights of the model directly indicate the significance of each gene and thus, allow a straightforward interpretation of the obtained results.

We demonstrated the usefulness of the approach using microarray datasets from four different disease areas, i.e., Psoriasis, Multiple Sclerosis, Chronic Obstructive Pulmonary Disease and Lung Cancer. For most of the prediction tasks, the classification algorithm performed reasonably well. In particular, the Psoriasis datasets were handled surprisingly well. In cases where weak scores were achieved, we performed additional analyses to

pinpoint the factors that may have lead to a decreased performance. For instance, in case of the MSD dataset, our results from the leave-one-out cross-validation study indicate significant qualitative differences between the training and test datasets. Our results demonstrate that statistical methods in conjunction with modern microarray gene expression technology provide powerful and important means to accurately diagnose complex diseases.

MATERIALS AND METHODS

Data normalization

For all sub-challenges, only training data stemming from Affymetrix GeneChip Human Genome U133 Plus 2.0 microarrays were used, since all test datasets were generated on this platform. This was done to avoid any bias to our model that could have been introduced by including data from other chips in the training phase. Thus, normalization between different types of microarray chips was not needed, but normalization to remove batch effects between different experiments was still essential to make the datasets comparable. We normalized the pooled datasets (comprising both the training and test datasets for each of the individual sub-challenges) using a standard RMA normalization procedure [106].

Feature pre-selection

Before we used the datasets to train the classifier, the dimensionality of the feature space was reduced substantially by applying a feature preselection method. The aim was to select only those features that were significantly up or down regulated between case and control groups. We applied a nonparametric method based on the Wilcoxon ranksum test statistic [51, 173]. For each feature, we tested the null hypothesis that the distributions of its expression value over the case and control probes in the microarray datasets are equal, against the alternative that one distribution is stochastically larger than the other. This test is equivalent to the Wilcoxon two-sample test (also known as the Mann-Whitney U test). For each gene g, we obtain,

$$\text{Score}\,(g) = \sum_{i \in N_0} \sum_{j \in N_1} 1_{\{x_j^{(g)} - x_i^{(g)}\}}, \tag{88}$$

where $x_j^{(g)}$ is the expression value of gene g for an individual i and N_m represents the set of indices having a response in $m \in \{0, 1\}$. The score function counts the number of instances where an expression value corresponding to a response 1 is smaller than an expression value corresponding to a response 0. Therefore, the score would be close to the maximum score $|N_0|\,|N_1|$ for any gene that tends to be under-expressed in response

1 and close to 0 for a gene that tends to be over-expressed in individuals in N_1.

Clearly, the aim was to identify genes with small p-values for the corresponding Wilcoxon two-sample test, which is based on the test statistic Score(g). At 10% level of significance, we selected only the genes that had p-values less than 0.1/(total number of genes), using the Bonferroni correction under the multiple comparison setup.

Although this method of pre-selection can filter out genes that are predictive individually, it does not help to identify the best predictive combination of genes. For this reason, if the resultant dataset contains very few genes with p-values less than 0.1/(total number of genes), the resultant dataset will no more be reliable since some valuable information might have already been thrown away. In addition, the Bonferroni significance level is quite conservative. To avoid an excessive loss of features, the first 2000 genes, ordered by their p-values were picked if the pre-selection method initially yielded less than 2000 genes.

Training the primary classifier

We used a logistic regression model to fit the training data and to classify the test data. Despite the feature pre-selection, the feature space was yet 4% - 28% of the total number of probes on the chip (54,675). This was still high compared to the training data sample size. A simple logistic regression model [44] would lead to overfitting [25]. We therefore used an L1-regularized logistic regression model to drive a large number of less significant parameters to 0 and filter out only those genes that played a significant role in classifying the data into case and control groups.

Let $Y_i \in \{0, 1\}$ be the random variable that represents the response of the ith individual. Now we define the standardized expression value of gene g for individual i by

$$z_i^{(g)} = \frac{x_i^{(g)} - \mu_g}{\sigma_g}, \tag{89}$$

with

$$\mu_g = \frac{1}{n} \sum_{i=1}^{n} x_i^{(g)} \tag{90}$$

and

$$\sigma_g = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i^{(g)} - \mu_g \right)^2. \tag{91}$$

Then our model is

$$\pi_i := \Pr(Y_i = 1) = \frac{\exp\left(\alpha + \sum_{g=1}^{p} \beta_g z_i^{(g)}\right)}{1 + \exp\left(\alpha + \sum_{g=1}^{p} \beta_g z_i^{(g)}\right)}, \tag{92}$$

where p is the total number of genes under consideration. Hence, the likelihood of the observed data is

$$\begin{aligned} L(\alpha, \beta_1, \ldots, \beta_p \mid y_1, \ldots, y_n) &:= Pr(Y_1 = y_1, \ldots, Y_n = y_n) \\ &= \prod_{i=1}^{n} \frac{\left[\exp(\alpha + \sum_{g=1}^{p} \beta_g z_i^{(g)}\right]^{1_{\{y_i=1\}}}}{1 + \exp(\alpha + \sum_{g=1}^{p} \beta_g z_i^{(g)})}. \end{aligned} \tag{93}$$

Therefore, an estimate of the parameter-vector $\theta = (\alpha, \beta_1, \ldots, \beta_p)'$ can be obtained by maximizing the log-likelihood function as

$$\begin{aligned} \hat{\theta} &= \underset{\theta}{\operatorname{argmax}} \log L(\alpha, \beta_1, \ldots, \beta_p \mid y_1, \ldots, y_n) \\ &= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \left[1_{\{y_i=1\}} \left(\alpha + \sum_{g=1}^{p} \beta_g z_i^{(g)} \right) \right. \\ &\quad \left. - \log \left(1 + \exp \left(\alpha + \sum_{g=1}^{p} \beta_g z_i^{(g)} \right) \right) \right]. \end{aligned} \tag{94}$$

As mentioned earlier, we had to avoid overfitting, and thus optimized a penalized log-likelihood with an L1 penalty in β_g as

$$J(\theta) = \log L(\alpha, \beta_1, \ldots, \beta_p \mid y_1, \ldots, y_n) + \lambda \left(|\alpha| + \sum_{g=1}^{p} |\beta_g| \right), \tag{95}$$

The regularization or tuning parameter λ was fixed to the value that yielded the lowest L1-regularized deviance ($-2J(\theta)$), out of a 30-fold cross-validation on the training dataset. Figure 54 shows the cross-validated deviance estimates and confidence bounds for each proposed λ, as well as the selection of the optimal regularization parameter for the LC (2 classes) task.

Note that this is a convex optimization problem that can be solved efficiently. We used the MATLAB *lassoglm()* function, which uses the coordinate descent algorithm [68] to solve the optimization problem for a given regularization parameter λ. After obtaining the estimates of the parameter vector, the probability that an individual with expression value $x^{(g)}$ for gene g, belongs to class 1 (i.e. has the response 1), is given by

$$\hat{\pi}\left(x^{(1)}, \ldots, x^{(p)}\right) = \frac{\exp\left(\hat{\alpha} + \sum_{g=1}^{p} \hat{\beta}_g \frac{x^{(g)} - \mu_g}{\sigma_g}\right)}{1 + \exp\left(\hat{\alpha} + \sum_{g=1}^{p} \hat{\beta}_g \frac{x^{(g)} - \mu_g}{\sigma_g}\right)}. \tag{96}$$

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

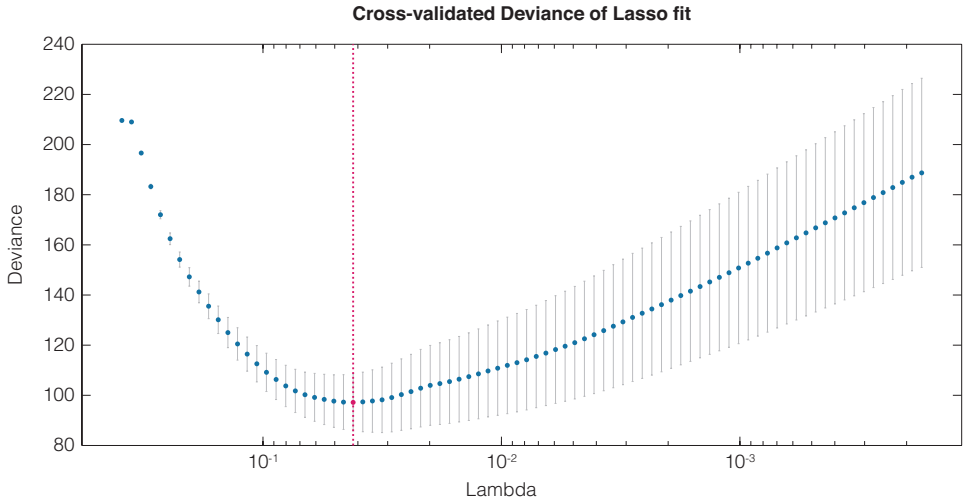

Figure 54: Cross-validated estimates of the deviance and confidence bound of the LASSO fit.

ACKNOWLEDGEMENTS

We thank the sbv IMPROVER committee for organizing the Diagnostic Signature Challenge. The challenge provided a great opportunity for us to develop our classification algorithms and validate it using real-life experimental datasets.

Part V

CONCLUSION

10 DISCUSSION

Research in biology is increasingly enhanced by data-driven methods. Available experimental techniques already allow to generate masses of data with a precision down to the single cell or molecule level, that can be used for learning quantitative models, or for drawing conclusions based on statistical approaches. Due to the incredible complexity of biology, however, the formulation of hypothesis is, and will probably remain an essential part in biological sciences. Experimental conditions and dynamic perturbations need to be precisely defined to produce data sufficient for computational methods to be capable of addressing specific hypothesis.

Several related questions were at the core of this thesis:

- How to use model-based design of experiments to generate data, optimized for specific modeling questions?
- How to implement and optimize dynamic experiments for live cell microscopy experiments?
- How to use data and computational tools to validate existing, or formulate new hypothesis?

Part ii addressed the design and implementation of dynamic perturbation experiments for inference tasks related to the modeling of stochastic reaction networks. I started with implementing microfluidic devices to synthesize temporal perturbation profiles during live cell fluorescence microscopy experiments. In Chapter 3, I introduced the general concept of PWM for liquid flows using an external 3-way valve and commercially available flow chambers (ibidi), and demonstrated its effectiveness by diluting profiles of fluorescent dye, and by maintaining the activation of the MAPK Hog1 by ramping of extracellular NaCl concentration. While the custom valve-on-chip PDMS chips, introduced in Chapter 7 have several advantages, the approach using the external valves and ibidi flow chambers can be used even with small hydrophobic molecules, making it compatible with several inhibitors or inducers that would be absorbed into PDMS, making it our setup of choice for flow experiments in Chapter 6. Further, the use of commercially available flow chambers readily allows the use of other cell types than yeast. I already tested the setup using human bone osteosarcoma epithelial cells from the U2OS cell line with great success[1].

Chapter 4 focused on Monte Carlo techniques to determine the expected information gain (e.g. in form of the KL divergence) between prior and posterior distributions of model parameters for proposed perturbation profiles. We presented an analytical proof for an optimal input sequence for

1 Data not shown in this thesis.

a simple Birth/Death process, given complete observations, and extended the framework for a more realistic scenario of incomplete and noisy observations. However, finding an optimal perturbation sequence using this approach remains challenging, as only a set of proposed input sequences can be assessed, and the high-dimensional Monte Carlo sampling remains computationally demanding and drastically limits the scalability of the problems under study. To increase the practical usability of the Bayesian optimal design framework, we formulated it as a variational problem in Chapter 5. A stochastic approximation scheme was used to numerically estimate the gradient of an objective function, and minimize it by iterative updates of a constrained perturbation sequence. Using an importance sampling scheme, the number of required SSA runs could be significantly reduced, making the framework applicable to automatically determine optimal perturbation sequences to maximally excite larger, and in combination with its ability to deal with incomplete and noisy measurements, practically more relevant model systems.

The contribution to Chapter 6, which introduced the Bayesian inference framework DPP, was an experimental case study system. We engineered an artificial gene expression system into a yeast strain, where upon addition of the exogenous hormone β-estradiol, the expression of a fluorescent reporter protein under a GAL1 promoter is induced. Recordings of fluorescence intensities of the expressed FP sVenus were mapped to absolute copy numbers of FP molecules by a specifically recorded calibration curve. To acquire this curve, we labeled several proteins of known cellular abundance in individual strains, recorded their fluorescent intensities using the same imaging conditions, and fitted a curve to the measured data points. To follow transcription events closer, we further added a destabilizing sequence to reduce the half-life of the expressed FP. We recorded the cellular response to pulses of β-estradiol, generated using a flow chamber setup as described earlier, and used the resulting relocation data of the specific TF GEV, acquired through a fusion of the TF to the FP mCherry, with the expression data as input/output data for a case study of the DPP framework. Validation experiments, using a dual reporter system were performed to assess the capability of the inference algorithm to dissect noise contributions into intrinsic and extrinsic components.

In Chapter 7 I summarized preliminary results of an attempt to combine the methods introduced in Chapter 3 - 6 to investigate a transcriptional network regulated by the MAPK Hog1. First I designed and manufactured a versatile PDMS chip that combines the synthesis of temporal perturbation profiles, using PWM, with single cell traps to capture individual yeast cells. The implementation as a PDMS device had several advantages compared to the valve-off-chip approach, I introduced earlier. The reduced flow rate and volume increased the maximal duration of an experiment, reduced shear stress on the cells, but most importantly the lower height of the cell chamber increased the imaging quality, while less movement of the cells helped to better maintain the acquisition focal plane.

In a dose response curve, acquired by flow cytometry, I showed how extracellular concentrations of NaCl induce the expression of pSTL1 driven qV expression, which shows a bimodal behavior across the population for intermediate NaCl concentrations. We refitted an ODE model to predict the Hog1-mCherry relocation upon temporal profiles of NaCl concentration, and used the relocation data in combination with calibrated single cell traces of qV expression to calibrate a three state gene expression model using the DPP framework. In a second inference iteration we compared the information gain between prior and posterior distributions of kinetic parameters that datasets recorded with input sequences of different complexity yielded. While the more complex sequences yielded a significantly higher information gain for rate estimates, compared to the sequence of less complexity, the effect of both complex sequences could hardly be distinguished. As single cell expression trajectories can not follow highly complex input sequences closely, one could assume that the majority of the transcription dynamics remains hidden below the slow dynamics of FP expression. Alternative approaches to follow transcription events closer could potentially resolve this issue. While the prediction of expected information gains for novel input stimuli is still ongoing, an application of the OED scheme introduced in Chapter 5 was so far limited by computational resources.

Chapter 8 of Part iii we formalized a hypothesis using a stochastic computational model, developed new hypothesis based on model predictions, and validated these experimentally. In the context of cell-cell communication, we investigated the sensing and decoding of spatial signals (i.e. chemical gradients). We could computationally define a generalizable mechanism how cells use gradient directions and intensities to align their growth axis towards the gradients source. For a specific yeast model system, we could test the predicted molecular mechanism and show that a mobile polarity site is essential to sense a gradient locally. By taking repeated directional decisions, the polarity site can filter fluctuations due to noise, and over time move towards the gradient direction.

In Part iv I presented our contribution to the sbv IMPROVER Diagnostic Signature Challenge, where we developed a classification pipeline to establish diagnostic signatures and predict instances of four disease areas (i.e. psoriasis, MS, COPD, LC). Our data-driven approach was centered around L_1-regularized logistic regression. While our approach scored reasonably well, the predictions were more accurate for some diseases than for others. This observation goes in hand with an analysis performed after the challenge was closed [223], which revealed that the quality of predictions depends more on a disease endpoint than on a particular classification approach. The prediction accuracy was significantly better for diseases where measurements were performed using primary tissue (e.g. skin in psoriasis, tumor cells in LC) instead of indirect targets such as blood in MS.

While computational tools, statistical methods and machine learning approaches for the reverse-engineering of intracellular processes become in-

creasingly effective, their success critically depends on the datasets available. As a common thread through Parts ii-iv, we could observe that all cases essentially depended on the precision, resolution and scope of experimental methods for perturbing and observing dynamic cellular processes. With the development of novel or refined experimental tools, new insights will be possible, and the advance of computational tools and computational power will further enhance possibilities. But the full potential will only be realized as both sides, experimental and computational strategies, are carefully matched.

11

OUTLOOK

In the ongoing project of Chapter 7, we are in the process of validating the inference results and run DPP iterations using datasets generated with various input perturbation sequences. Reconstructions of unobserved molecular states have so far looked promising, but the simulation of new cell trajectories has remained challenging. An efficient optimization framework should be applied to automatically determine optimal perturbation sequences for maximizing the information gain with respect to the inference of model parameters, or for specific model selection tasks. I intend to address questions regarding the structure and kinetic properties of the transcriptional response network to osmotic stress. While the promoters controlled by the MAPK Hog1 have differences in structure, transcription factors involved or their expression dynamics, a systematic comparison of model hypothesis would be a valuable case study for optimal model selection tasks.

To overcome the limitations of slow FP reporter dynamics, alternative approaches to directly record transcription output could be ideally suited for inference tasks of transcription networks. Using the MS2 or PP7 system, fast acquisition bursts could be applied to estimate the current rate of mRNA production and abundance, while preserving the cell viability and limiting the influence of phototoxicity. The distribution of these acquisition burst time points, as well as other measurement time points could be another optimization criterion and be jointly optimized with an input perturbation sequence in the OED framework.

Further, investigations into optimization of several simultaneous experiments could be performed. In this case, one would intuitively expect the OED framework to yield perturbation sequences of different dynamics, possibly focusing on different parts of the underlying dynamic system.

The microfluidic devices presented in this thesis can be readily applied, or easily adapted to multiple other tasks of perturbing cellular states with temporal concentration profiles and be used in various experimental settings.

Part VI

APPENDIX

A

FABRICATION OF MICROFLUIDIC DEVICES

A.1 PRODUCTION OF WAFER MOLDS

Microfluidic chips are created from two silicon wafer molds produced by soft lithography [249]. The *flow* wafer serves as a mold for flow channels that consist of four layers: Cell Trap Chamber, Cell Loading Channels, Media Outlet and Media Inlet. The second, *control*, wafer serves as a mold for the control channels for opening/closing the the media inlets.

Fabrication conditions were chosen according to the manual, provided by the manufacturer of the photoresists (MicroChem). Rounded, in cross-section view, shapes of the media inlet channels were fabricated using positive (AZ) photoresist [233]. All other structures are fabricated using negative photoresist (SU 8) for a rectangular shape in the cross-sectional view. Before spreading photoresist, bare wafers were dehydrated to enhance adhesion by placing them on a hot plate with 130° C for 15 min. The UV exposure and alignment of wafers were done with a mask aligner (MA6, Karl Süss). The total energy levels of exposing UV light were calculated based on intensity measurements at 365 nm for SU 8 and at 405 nm for AZ. After fabrication of each layer, the structures were examined by microscopy, and the operation condition were modified accordingly (e.g. increasing the UV exposure or developing time). Further, each layer fabricated layer was covered using adhesive tape (Scotch-tape, 3M) to prevent spreading of the photoresist on the alignment marks. This step has shown to be especially important for fabricating the cell trap chamber, and structures were gently covered by tape during the spinning of the AZ photoresist. The detailed operation conditions are summarized in Table 19.

LAYER	CELL TRAP CHAMBER	CELL LOADING CHANNELS	MEDIA OUTLET	MEDIA INLET
Photo Resist	SU-8 2005	SU-8 10	SU-8 2025	AZ 4562
Aimed Height [µm]	5	15	25	10
Spinning Speed [rpm]	3000	2000	3000	1700
Soft Bake 1 (65° C) [min]	x	2	x	x
Soft Bake 2 (95° C) [min]	2	5	6	10
Exposure [$J\,cm^{-2}$]	120	280	200	700
Post Exposure Bake 1 (65° C) [min]	x	1	x	x
Post Exposure Bake 2 (95° C) [min]	3	2	6	x
Developing Time [min]	1	3	4	6
Reflow Time (130° C) [min]	x	x	x	15

Table 19: Operating conditions for the production of silicon wafers for the PWM/cell trap chip.

A.2 PRODUCTION OF PDMS CHIPS

STEP 1 Coat wafer mold surface. Put silicone wafers into an exsiccator, add 80 µl of 1H,1H,2H,2H-Perfluorodecyltrichlorosilane (abcr) in a porcelaine dish and apply vacuum, incubate for > 24 h; This step is only required once. *NOTE: This step has to be performed in a fume hood.*

STEP 2 Control Layer Mixture Ratio: 5:1 (35 g of Sylgard 184 A + 7 g of Sylgard 184 B); Mix both components, degas in vacuum chamber, and pour onto *control* mold. Degas again.

STEP 3 Flow Layer Mixture Ratio: 20:1 (10 g of Sylgard 184 A + 1 g of Sylgard 184 B); Mix both components, degas, and spin coat onto *flow* mold at 1700 rpm for 40 s.

STEP 4 Place the both molds into an oven, set to 80° and incubate for 30 min.

STEP 5 Remove molds from oven, remove PDMS from the *control* mold, cut out individual chips, punch holes for valve inlets, and align to flow layer.

STEP 6 Put the aligned device back into the oven, set to 80° and incubate for > 120 min.

STEP 7 Remove devices from oven, cut out chips, and punch *flow* layer holes.

STEP 8 Bond PDMS chips to glass coverslips using UV treatment.

STEP 9 Place the freshly bonded chips into an oven, set to 80° and incubate for 15 min.

BIBLIOGRAPHY

[1] O Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–26, 1978.

[2] O Aalen. Mixing distributions on a Markov Chain. *Scandinavian Journal of Statistics*, 14(4):281–9, 1987.

[3] O Aalen, O Borgan, and H Gjessing. *Survival and event history analysis: A process point of view*. Springer, 2008.

[4] M Acar, J T Mettetal, and A van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics*, 40(4):471–5, April 2008.

[5] C E Adler, R D Fetter, and C I Bargmann. UNC-6/Netrin induces neuronal asymmetry and defines the site of axon formation. *Nature Neuroscience*, 9(4):511–8, April 2006.

[6] A Ainla, I Goezen, O Orwar, and A Jesorka. A microfluidic diluter based on pulse width flow modulation. *Analytical Chemistry*, 81(13):5549–56, July 2009.

[7] S J Altschuler and L F Wu. Cellular heterogeneity: Do differences make a difference? *Cell*, 141(4):559–63, May 2010.

[8] M Amrein and H R Kuensch. Rate estimation in partially observed Markov jump processes with measurement errors. *Statistics and Computing*, 22(2):513–26, March 2011.

[9] D F Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of Chemical Physics*, 127:214107, December 2007.

[10] D F Anderson and T G Kurtz. Continuous time Markov Chain Models for chemical reaction networks. In H Koeppl, G Setti, M di Bernardo, and D Densmore, editors, *Design and Analysis of Biomolecular Circuits*, pages 3–42. Springer New York, 2011.

[11] N Andrew, D Craig, J P Urbanski, J Gunawardena, and T Thorsen. Microfluidic temporal cell stimulation. In *12th International Conference on Miniaturized Systems for Chemistry and Life Sciences*, San Diego, CA, USA, 2008. CBMS.

[12] J F Apgar, J E Toettcher, D Endy, F M White, and B Tidor. Stimulus design for model selection and validation in cell signaling. *PLoS Computational Biology*, 4(2):e30, February 2008.

[13] M Ashyraliyev, Y Fomekong-Nanfack, J A Kaandorp, and J G Blom. Systems biology: parameter estimation for biochemical models. *The FEBS journal*, 276(4):886–902, February 2009.

[14] K R Ayscough and D G Drubin. A role for the yeast actin cytoskeleton in pheromone receptor clustering and signalling. *Current Biology*, 8(16):927–31, July 1998.

[15] J Baehler, J Q Wu, M S Longtine, N G Shah, A McKenzie, A B Steever, A Wach, P Philippsen, and J R Pringle. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast*, 14(10):943–51, July 1998.

[16] R Bahar, C H Hartmann, K A Rodriguez, A D Denny, R A Busuttil, M E T Dolle, R B Calder, G B Chisholm, B H Pollock, C A Klein, and J Vijg. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, 441(7096):1011–4, June 2006.

[17] N Q Balaban, J Merrin, R Chait, L Kowalik, and S Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–5, September 2004.

[18] S Bandara, J P Schloeder, R Eils, H G Bock, and T Meyer. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Computational Biology*, 5(11):e1000558, November 2009.

[19] D R Bandura, V I Baranov, O I Ornatsky, A Antonov, R Kinach, X Lou, S Pavlov, S Vorobiev, J E Dick, and S D Tanner. Mass cytometry: Technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Analytical Chemistry*, 81(16):6813–22, August 2009.

[20] E Batchelor, A Loewer, C Mock, and G Lahav. Stimulus-dependent dynamics of p53 in single cells. *Molecular Systems Biology*, 7(488), May 2011.

[21] S C Bendall, E F Simonds, P Qiu, E D Amir, P O Krutzik, R Finck, RV Bruggner, R Melamed, A Trejo, O I Ornatsky, R S Balderas, S K Plevritis, K Sachs, D Pe'er, S D Tanner, and G P Nolan. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*, 332(6030):687–96, May 2011.

[22] S Bersini, J S Jeon, G Dubini, C Arrigoni, S Chung, J L Charest, M Moretti, and R D Kamm. A microfluidic 3D in vitro model for specificity of breast cancer metastasis to bone. *Biomaterials*, 35(8):2454–61, March 2014.

[23] E Bertrand, P Chartrand, M Schaefer, S M Shenoy, R H Singer, and R M Long. Localization of ASH1 mRNA particles in living yeast. *Molecular Cell*, 2(4):437–45, October 1998.

[24] A Berzat and A Hall. Cellular responses to extracellular guidance cues. *The EMBO Journal*, 29(16):2734–45, August 2010.

[25] C M Bishop. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[26] W J Blake, M Kaern, C R Cantor, and J J Collins. Noise in eukaryotic gene expression. *Nature*, 422:633–7, April 2003.

[27] M Blondel, P M Alepuz, L S Huang, S Shaham, G Ammerer, and M Peter. Nuclear export of Far1p in response to pheromones requires the export receptor Msn5p / Ste21p. 13:2284–300, 1999.

[28] I Bose, J E Irazoqui, J J Moskow, E S Bardes, T R Zyla, and D J Lew. Assembly of scaffold-mediated complexes containing Cdc42p, the exchange factor Cdc24p, and the effector Cla4p required for cell cycle-regulated phosphorylation of Cdc24p. *The Journal of Biological Chemistry*, 276(10):7176–86, March 2001.

[29] C G Bowsher and P S Swain. Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences of the United States of America*, May 2012.

[30] C B Brachmann, A Davies, G J Cost, E Caputo, J Li, P Hieter, and J D Boeke. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, 14:115–132, 1998.

[31] O Brandman and T Meyer. Feedback Loops Shape Cellular Signals in Space and Time. *Science*, 322:390–5, 2008.

[32] M P S Brown, W N Grundy, D Lin, N Cristianini, C W Sugnet, T S Furey, M Ares, and D Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences of the United States of America*, 97(1):262–7, January 2000.

[33] A G Busetto, A Hauser, G Krummenacher, M Sunnaker, S Dimopoulos, C S Ong, J Stelling, and J M Buhmann. Near-optimal experimental design for model selection in systems biology. *Bioinformatics*, 29(20):2625–32, October 2013.

[34] A Butty, P M Pryciak, L S Huang, I Herskowitz, and M Peter. The Role of Far1p in Linking the Heterotrimeric G Protein to Polarity Establishment Proteins During Yeast Mating. *Science*, 282(5393):1511–1516, November 1998.

[35] A C Butty, N Perrinjaquet, A Petit, M Jaquenoud, J E Segall, K Hofmann, C Zwahlen, and M Peter. A positive feedback loop stabilizes the guanine-nucleotide exchange factor Cdc24 at sites of polarization. *The EMBO Journal*, 21(7):1565–1576, 2002.

[36] Y Cao, D T Gillespie, and L R Petzold. Avoiding negative populations in explicit Poisson tau-leaping. *The Journal of chemical physics*, 123(5):054104, August 2005.

[37] A P Capaldi, T Kaplan, Y Liu, N Habib, A Regev, N Friedman, and E K O'Shea. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nature Genetics*, 40(11):1300–6, November 2008.

[38] M Chalfie, Y Tu, G Euskirchen, W W Ward, and DC Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263:1766–7, October 1994.

[39] K Chaloner and I Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

[40] R E Chen and J Thorner. Function and regulation in MAPK signaling pathways: Lessons learned from the yeast Saccharomyces cerevisiae. *Biochimica et biophysica acta*, 1773(8):1311–1340, 2007.

[41] J Chenevert, N Valtz, and I Herskowitz. Identification of genes required for normal pheromone-induced cell polarization in Saccharomyces cerevisiae. *Genetics*, pages 1287–97, 1994.

[42] C D Chin, T Laksanasopin, Y K Cheung, D Steinmiller, V Linder, H Parsa, J Wang, H Moore, R Rouse, G Umviligihozo, E Karita, L Mwambarangwe, S L Braunstein, J van de Wijgert, R Sahabo, J E Justman, W El-Sadr, and S K Sia. Microfluidics-based diagnostics of infectious diseases in the developing world. *Nature Medicine*, 17(8):1015–9, August 2011.

[43] A A Cohen, N Geva-Zatorsky, E Eden, M Frenkel-Morgenstern, I Issaeva, A Sigal, R Milo, C Cohen-Saidon, Y Liron, Z Kam, L Cohen, T Danon, N Perzov, and U Alon. Dynamic proteomics of individual cancer cells in response to a drug. *Science*, 322(5907):1511–6, December 2008.

[44] D Collett. *Modelling Binary Data*. Chapman and Hall/CRC, 2nd edition, 2003.

[45] A Colman-Lerner, A Gordon, E Serra, Ti Chin, O Resnekov, D Endy, C G Pesce, and R Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437:699–706, September 2005.

[46] M M Crane, I B N Clark, E Bakker, S Smith, and P S Swain. A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast. *PloS one*, 9(6):e100042, January 2014.

[47] S Crosson, S Rajagopal, and K Moffat. The LOV domain family: Photoresponsive signaling modules coupled to diverse output domains. *Biochemistry*, 42(1):2–10, 2003.

[48] I Dang, R Gorelik, C Sousa-Blin, E Derivery, C Guerin, J Linkner, M Nemethova, J G Dumortier, F A Giger, T A Chipysheva, V D Ermilova, S Vacher, V Campanacci, I Herrada, A-G Planson, S Fetics, V Henriot, V David, K Oguievetskaia, G Lakisic, F Pierre, A Steffen, A Boyreau, N Peyrieras, K Rottner, S Zinn-Justin, J Cherfils, I Bieche, A Y Alexandrova, N B David, J V Small, J Faix, L Blanchoin, and A Gautreau. Inhibitory signalling to the Arp2/3 complex steers cell migration. *Nature*, 503(7475):281–4, November 2013.

[49] K M Dean and A E Palmer. Advances in fluorescence labeling strategies for dynamic cellular imaging. *Nature Chemical Biology*, 10(7):512–23, June 2014.

[50] M Dettling and P Buehlmann. Finding predictive gene groups from microarray data. *Journal of Multivariate Analysis*, 90:106–31, July 2004.

[51] M Dettling and P Buhlmann. Boosting for tumor classification with gene expression data. *Bioinformatics*, 19(9):1061–9, June 2003.

[52] C G Dotti, C A Sullivan, and G A Banker. The Establishment of polarity by hippocampal neurons in culture. *The Journal of Neuroscience*, 8(4):1454–68, 1988.

[53] A Doucet, N de Freitas, and N Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[54] A Doucet, N de Freitas, K Murphy, and S Russell. Rao-Blackwellised particle filtering for dynamic bayesian networks. In *16th Conference on Uncertainty in Artificial Intelligence*, pages 176–83, Stanford, CA, USA, 2000.

[55] S Dudoit, J Fridlyand, and T P Speed. Comparison of discrimination methods for the classification of tumors using gene expresion data. *Journal of the American Statistical Association*, 97(457):77–87, 2002.

[56] J M Dyer, N S Savage, M Jin, T R Zyla, T C Elston, and D J Lew. Tracking shallow chemical gradients by actin-driven wandering of the polarization site. *Current Biology*, 23(1):32–41, January 2013.

[57] H Easwaran, H C Tsai, and S B Baylin. Cancer epigenetics: Tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Molecular Cell*, 54(5):716–727, June 2014.

[58] A Edelstein, N Amodaj, K Hoover, R Vale, and N Stuurman. Computer control of microscopes using MicroManager. *Current Protocols in Molecular Biology*, 92, October 2010.

[59] A Eldar and M B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–73, September 2010.

[60] J Elf and M Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research*, 13(11):2475–84, November 2003.

[61] M B Elowitz, A J Levine, E D Siggia, and P S Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–6, August 2002.

[62] E L Elson. Fluorescence correlation spectroscopy: past, present, future. *Biophysical Journal*, 101(12):2855–70, December 2011.

[63] Fedorov V V and P Hackl. *Model-oriented design of experiments (Lecture notes in statistics)*. Springer, 1997.

[64] A M Femino, F S Fay, K Fogarty, and R H Singer. Visualization of single RNA transcripts in situ. *Science*, 280:585–590, April 1998.

[65] J E Jr Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in Chemical Biology*, 14(2):140–148, April 2002.

[66] P Flaherty, M Jordan, and A P Arkin. Robust design of biological experiments. In *Advances in Neural Information Processing Systems 18*, pages 363–370. MIT Press, 2006.

[67] J Friedman, T Hastie, and R Tibshirani. *The elements of statistical learning: Data mining, inference, and prediction*. Springer, 2nd edition, 2009.

[68] J Friedman, T Hastie, and R Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

[69] N Friedman, L Cai, and X S Xie. Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression. *Physical Review Letters*, 97:168302, October 2006.

[70] G F Froment, K B Bischoff, and J De Wilde. *Chemical reactor analysis and design*. John Wiley and Sons, 3rd edition, 2010.

[71] T S Furey, N Cristianini, N Duffy, W David, D W Bednasrki, M Schummer, and D Haussler. Support vector machine classification and validation of cancer tissues samples using microarray expression data. 16:906–914, 2000.

[72] A Gartner, A Jovanovic, D I Jeoung, S Bourlat, F R Cross, and G Ammerer. Pheromone-dependent G1 cell cycle arrest requires Far1 phosphorylation, but may not involve inhibition of Cdc28-Cln2 kinase, In vivo. *Molecular and Cellular Biology*, 18(7):3681–91, 1998.

[73] A Gautier, C Gauron, M Volovitch, D Bensimon, L Jullien, and S Vriz. How to control proteins with light in living systems. *Nature Chemical Biology*, 10(7):533–41, June 2014.

[74] A Gautier, A Juillerat, C Heinis, I R Correa, M Kindermann, F Beaufils, and K Johnsson. An engineered protein tag for multiprotein labeling in living cells. *Chemistry and Biology*, 15(2):128–36, February 2008.

[75] S Ghaemmaghami, W K Huh, K Bower, R W Howson, A Belle, N Dephoure, E K O'Shea, and J S Weissman. Global analysis of protein expression in yeast. *Nature*, 425:737–41, October 2003.

[76] D T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–34, 1976.

[77] D T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–61, 1977.

[78] D T Gillespie. *Markov Processes: An introduction for physical scientists*. Academic Press, Inc., New York, 1992.

[79] D T Gillespie. The multivariate Langevin and Fokker-Planck equations. *American Journal of Physics*, 64(10):1246, 1996.

[80] D T Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[81] D T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716, 2001.

[82] D T Gillespie and L R Petzold. Improved leap-size selection for accelerated stochastic simulation. *The Journal of Chemical Physics*, 119(16):8229, 2003.

[83] A Golightly and D J Wilkinson. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, 61:781–8, September 2005.

[84] A Golightly and D J Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, 1(6):807–820, September 2011.

[85] D K Gonda, A Bachmair, I Wuenning, J W Tobias, W S Lane, and A Varshavsky. Universality and structure of the N-end rule. *The Journal of Biological Chemistry*, 264(28):16700–12, October 1989.

[86] A B Goryachev and A V Pokhilko. Computational model explains high activity and rapid cycling of Rho GTPases within protein complexes. *PLoS Computational Biology*, 2(12):e172, December 2006.

[87] M Gossen and H Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12):5547–51, June 1992.

[88] B R Graziano and O D Weiner. Self-organization of protrusions and polarity during eukaryotic chemotaxis. *Current Opinion in Cell Biology*, 30:60–7, October 2014.

[89] E A Hackett, R K Esch, S Maleri, and B Errede. A family of destabilized cyan fluorescent proteins as transcriptional reporters in S. cerevisiae. *Yeast*, 23:333–49, April 2006.

[90] N Hao, B Budnik, J Gunawardena, and E K O'Shea. Tunable signal processing through modular control of transcription factor translocation. *Science*, 339(6118):460–4, January 2013.

[91] Nan Hao and Erin K O'Shea. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nature Structural and Molecular Biology*, 19(1):31–9, January 2012.

[92] C V Harper, B Finkenstaedt, D J Woodcock, S Friedrichsen, S Semprini, L Ashall, D G Spiller, J J Mullins, D A Rand, J R E Davis, and M R H White. Dynamic analysis of stochastic transcription cycles. *PLoS Biology*, 9(4):e1000607, April 2011.

[93] Y Hart and U Alon. The utility of paradoxical components in biological circuits. *Molecular Cell*, 49(2):213–21, January 2013.

[94] J Hasenauer, S Waldherr, M Doszczak, N Radde, P Scheurich, and F Allgoewer. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics*, 12:125, January 2011.

[95] F Hayot and C Jayaprakash. The linear noise approximation for molecular fluctuations within cells. *Physical Biology*, 1:205–10, December 2004.

[96] P Hersen, M N McClean, L Mahadevan, and S Ramanathan. Signal processing by the HOG MAP kinase pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 105(20):7165–70, May 2008.

[97] J Hespanha. Moment closure for biochemical networks. In *3rd International Symposium on Communications, Control and Signal Processing*, number March, pages 142–7, St. Julians, Malta, 2008. IEEE.

[98] L Hicke and H Riezman. Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell*, 84(2):277–87, January 1996.

[99] A Hilfinger and J Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29):12167–72, July 2011.

[100] M J Hinner and K Johnsson. How to obtain labeled proteins and what to do with them. *Current Opinion in Biotechnology*, 21(6):766–76, December 2010.

[101] D G Holmes and T A Lipo. *Pulse width modulation for power converters - Principles and practice*. IEEE/Wiley, 2003.

[102] J R Houser, E Ford, S M Chatterjae, S Maleri, T C Elston, B Errede, and S M Chatterjea. An improved short-lived fluorescent protein transcriptional reporter for Saccharomyces cerevisiae. *Yeast*, 29(12):519–30, October 2012.

[103] D Huh, B D Matthews, A Mammoto, M Montoya-Zavala, H Y Hsin, and D E Ingber. Reconstituting organ-level lung functions on a chip. *Science*, 328(5986):1662–8, June 2010.

[104] W K Huh, J V Falvo, L C Gerke, A S Carroll, R W Howson, J S Weissman, and E K O'Shea. Global analysis of protein localization in budding yeast. *Nature*, 425:686–91, October 2003.

[105] J E Irazoqui, A A Gladfelter, and D J Lew. Scaffold-mediated symmetry breaking by Cdc42p. *Nature Cell Biology*, 5(12):1062–70, December 2003.

[106] R A Irizarry, B Hobbs, F Collin, Y D Beazer-Barclay, K J Antonellis, U Scherf, and T P Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–64, April 2003.

[107] C Janetopoulos, L Ma, P N Devreotes, and P A Iglesias. Chemoattractant-induced phosphatidylinositol 3,4,5-trisphosphate accumulation is spatially amplified and adapts, independent of the actin cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America*, 101(24):8951–6, 2004.

[108] C Janke, M M Magiera, N Rathfelder, C Taxis, S Reber, H Maekawa, A Moreno-Borchart, G Doenges, E Schwob, E Schiebel, and M Knop. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast*, 21(11):947–62, August 2004.

[109] E A Jares-Erijman and T M Jovin. FRET imaging. *Nature Biotechnology*, 21(11):1387–95, November 2003.

[110] D D Jenness, A C Burkholder, and L H Hartwell. Binding of alpha-factor pheromone to Saccharomyces cerevisiae a cells: Dissociation constant and number of binding sites. *Molecular and Cellular Biology*, 6(1):318–20, 1986.

[111] D D Jenness and P Spatrick. Down regulation of the alpha-factor pheromone receptor in S. cerevisiae. *Cell*, 46(3):345–353, August 1986.

[112] A Jilkine, S B Angenent, L F Wu, and S J Altschuler. A density-dependent switch drives stochastic clustering and polarization of signaling molecules. *PLoS Computational Biology*, 7(11):e1002271, November 2011.

[113] A Keppler, S Gendreizig, T Gronemeyer, H Pick, H Vogel, and K Johnsson. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology*, 21(1):86–9, January 2003.

[114] J Khan, J S Wei, M Ringner, L H Saal, M Ladanyi, F Westermann, F Berthold, M Schwab, C R Antonescu, C Peterson, and P S Meltzer. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7(6):673–9, June 2001.

[115] J Kiefer and J Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–6, 1952.

[116] H Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–4, March 2002.

[117] M Knop, K Siegers, G Pereira, W Zachariae, B Winsor, K Nasmyth, and E Schiebel. Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast*, 15:963–72, July 1999.

[118] H Koeppl, C Zechner, A Ganguly, S Pelet, and M Peter. Accounting for extrinsic variability in the estimation of stochastic rate constants. *International Journal of Robust and Nonlinear Control*, 22(10):1103–19, 2012.

[119] D Koller and N Friedman. *Probabilistic graphical models: Principles and techniques*. MIT Press, 2009.

[120] M Komorowski, M J Costa, D A Rand, and M P H Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proceedings of the National Academy of Sciences of the United States of America*, 108(21):8645–50, May 2011.

[121] A Kramer, J Hasenauer, F Allgoewer, and N Radde. Computation of the posterior entropy in a Bayesian framework for parameter estimation in biological networks. In *IEEE International Conference on Control Applications*, volume 4, pages 493–8. IEEE, 2010.

[122] C Kreutz, M M Bartolome Rodriguez, T Maiwald, M Seidl, H E Blum, L Mohr, and J Timmer. An error model for protein quantification. *Bioinformatics*, 23(20):2747–53, October 2007.

[123] C Kreutz and J Timmer. Systems biology: experimental design. *The FEBS journal*, 276:923–42, February 2009.

[124] U Kuechler and M Sorensen. Exponential families of stochastic processes. Springer, 1997.

[125] S Kullback. *Information theory and statistics*. Dover Publications, 1997.

[126] H J Kushner. A projected stochastic approximation method for adaptive filters and identifiers. *IEEE Transactions on Automatic Control*, 25(4):836–8, 1980.

[127] D R Larson, C Fritzsch, L Sun, X Meng, D S Lawrence, and R H Singer. Direct observation of frequency modulated transcription in single cells using light activation. *eLife*, 2:e00750, January 2013.

[128] D R Larson, R H Singer, and D Zenklusen. A single molecule view of gene expression. *Trends in Cell Biology*, 19(11):630–7, November 2009.

[129] D R Larson, D Zenklusen, B Wu, J A Chao, and R H Singer. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*, 332(6028):475–8, April 2011.

[130] K E Lee, N Sha, E R Dougherty, M Vannucci, and B K Mallick. Gene selection: a Bayesian variable selection approach. *Bioinformatics*, 19(1):90–7, January 2003.

[131] S Lee, W A Lim, and K S Thorn. Improved blue, green, and red fluorescent protein tagging vectors for S. cerevisiae. *PloS one*, 8(7):e67902, January 2013.

[132] S S Lee, P Horvath, S Pelet, B Hegemann, L P Lee, and M Peter. Quantitative and dynamic assay of single cell chemotaxis. *Integrative Biology*, 4:381–90, April 2012.

[133] A Levskaya, O D Weiner, W A Lim, and C A Voigt. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature*, 461(7266):997–1001, October 2009.

[134] X Li, R J Quigg, J Zhou, W Gu, P N Rao, and E F Reed. Clinical utility of microarrays: current status, existing challenges and future outlook. *Current Genomics*, 9(7):466–74, November 2008.

[135] J W Lichtman and J A Conchello. Fluorescence microscopy. *Nature Methods*, 2(12):910–9, December 2005.

[136] J Liepe, S Filippi, M Komorowski, and M P H Stumpf. Maximizing the information content of experiments in systems biology. *PLoS Computational Biology*, 9(1):e1002888, January 2013.

[137] J Liepe, P Kirk, S Filippi, T Toni, C P Barnes, and M P H Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9(2):439–56, February 2014.

[138] D V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, December 1956.

[139] J Lippincott-Schwartz, E Snapp, and A Kenworthy. Studying protein dynamics in living cells. *Nature Reviews. Molecular Cell Biology*, 2(6):444–56, June 2001.

[140] A Loewer and G Lahav. We are all individuals: causes and consequences of non-genetic heterogeneity in mammalian cells. *Current Opinion in Genetics and Development*, 21(6):753–8, December 2011.

[141] G V Los, L P Encell, M G McDougall, D D Hartzell, N Karassina, C Zimprich, M G Wood, R Learish, R F Ohana, M Urh, D Simpson, J Mendez, K Zimmerman, P Otto, G Vidugiris, J Zhu, A Darzins, D H Klaubert, R F Bulleit, and K V Wood. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology*, 3(6):373–82, June 2008.

[142] J F Louvion, B Havaux-Copf, and D Picard. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene*, 131:129–34, September 1993.

[143] X Mao and T J Huang. Microfluidic diagnostics for the developing world. *Lab on a Chip*, 12(8):1412–6, April 2012.

[144] MAQC-Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology*, 24(9):1151–61, September 2006.

[145] D Marbach, J C Costello, R Kueffner, N M Vega, R J Prill, D M Camacho, K R Allison, The Dream 5 Consortium, M Kellis, J J Collins, and G Stolovitzky. Wisdom of crowds for robust gene network inference. 9(8), 2012.

[146] E Marco, R Wedlich-Soldner, R Li, S J Altschuler, and L F Wu. Endocytosis optimizes the dynamic localization of membrane proteins that regulate cortical polarity. *Cell*, 129(2):411–22, April 2007.

[147] A W Martinez, S T Phillips, G M Whitesides, and E Carrilho. Diagnostics for the developing world: microfluidic paper-based analytical devices. *Analytical Chemistry*, 82(1):3–10, January 2010.

[148] P B Mason and K Struhl. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Molecular Cell*, 17:831–40, March 2005.

[149] R S McIsaac, S J Silverman, M N McClean, P A Gibney, J Macinskas, M J Hickman, A A Petti, and D Botstein. Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. *Molecular Biology of the Cell*, 22:4447–59, November 2011.

[150] M A McMurray and J Thorner. Septin stability and recycling during dynamic structural transitions in cell division and development. *Current Biology*, 18(16):1203–8, August 2008.

[151] F Menolascina, D Bellomo, T Maiwald, V Bevilacqua, C Ciminelli, A Paradiso, and S Tommasi. Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering. *BMC bioinformatics*, 10:S4, January 2009.

[152] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller, and E Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–92, 1953.

[153] J T Mettetal, D Muzzey, C Gomez-Uribe, and A van Oudenaarden. The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science*, 319:482–4, 2008.

[154] P Meyer, L G Alexopoulos, T Bonk, A Califano, C R Cho, A de la Fuente, D de Graaf, A J Hartemink, J Hoeng, N V Ivanov, H Koeppl, R Linding, D Marbach, R Norel, M C Peitsch, J J Rice, A Royyuru, F Schacherer, J Sprengel, K Stolle, D Vitkup, and G Stolovitzky. Verification of systems biology research in the age of collaborative competition. *Nature Biotechnology*, 29(9):811–5, September 2011.

[155] P Meyer, J Hoeng, J J Rice, R Norel, J Sprengel, K Stolle, T Bonk, S Corthesy, A Royyuru, M C Peitsch, and G Stolovitzky. Industrial methodology for process verification in research (IMPROVER): toward systems biology verification. *Bioinformatics*, 28(9):1193–201, May 2012.

[156] C Michelot. A finite algorithm for finding the projection of a point onto the Canonical simplex or Rn. *Journal of Optimization Theory and Applications*, 50(1):1985–200, 1986.

[157] A Milias-Argeitis, S Summers, J Stewart-Ornstein, I Zuleta, D Pincus, H El-Samad, M Khammash, and J Lygeros. In silico feedback for in vivo regulation of a gene expression circuit. *Nature Biotechnology*, 29(12):1114–6, December 2011.

[158] C G Moles, P Mendes, and J R Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13(11):2467–74, November 2003.

[159] B Munsky and M Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124(4):044104, January 2006.

[160] B Munsky, B Trinh, and M Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(318), January 2009.

[161] D Muzzey, C A Gomez-Uribe, J T Mettetal, and A van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1):160–71, July 2009.

[162] P Nandy, M Unger, C Zechner, and H Koeppl. Optimal perturbations for the identification of stochastic reaction dynamics. In *16th IFAC Symposium on System Identification*, pages 686–91, Brussels, Belgium, July 2012. IFAC.

[163] A Nern and R A Arkowitz. A Cdc24p-Far1p-G-Beta-Gamma Protein Complex Required for Yeast Orientation during Mating. *The Journal of Cell Biology*, 144(6):1187–1202, 1999.

[164] A Nern and R A Arkowitz. Nucleocytoplasmic shuttling of the Cdc42p exchange factor Cdc24p. *The Journal of Cell Biology*, 148(6):1115–22, 2000.

[165] G Neuert, B Munsky, R Z Tan, L Teytelman, M Khammash, and A van Oudenaarden. Systematic identification of signal-activated stochastic gene regulation. *Science*, 339:584–7, February 2013.

[166] J R S Newman, S Ghaemmaghami, J Ihmels, D K Breslow, M Noble, J L DeRisi, and J S Weissman. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840–6, June 2006.

[167] R H Newman, M D Fosbrink, and J Zhang. Genetically encodable fluorescent biosensors for tracking signaling dynamics in living cells. *Chemical Reviews*, 111(5):3614–66, May 2011.

[168] J M K Ng, I Gitlin, A D Stroock, and G M Whitesides. Components for integrated poly(dimethylsiloxane) microfluidic systems. *Electrophoresis*, 23:3461–73, 2002.

[169] P R O'Neill and N Gautam. Subcellular optogenetic inhibition of G proteins generates signaling gradients and cell migration. *Molecular Biology of the Cell*, 25(15):2305–14, August 2014.

[170] R Opgen-Rhein and K Strimmer. Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics*, 8:S3, January 2007.

[171] M Opper and G Sanguinetti. Variational inference for Markov jump processes. In J C Platt, D Koller, Y Singer, and D Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, 2007.

[172] O Ornatsky, D Bandura, V Baranov, M Nitz, M A Winnik, and S Tanner. Highly multiparametric analysis by mass cytometry. *Journal of Immunological Methods*, 361:1–20, September 2010.

[173] P J Park, M Pagano, and M Bonetti. A nonparametric scoring algorithm for identifying informative genes from microarray data. In *Pacific Symposium on Biocomputing*, volume 63, pages 52–63, 2001.

[174] P Paszek, S Ryan, L Ashall, K Sillitoe, C V Harper, D G Spiller, D A Rand, and M R H White. Population robustness arising from cellular heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25):11644–9, June 2010.

[175] S Pelet, R Dechant, S S Lee, F van Drogen, and M Peter. An integrated image analysis platform to quantify signal transduction in single cells. *Integrative Biology*, 4:1274–82, October 2012.

[176] S Pelet, F Rudolf, M Nadal-Ribelles, E de Nadal, F Posas, and M Peter. Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science*, 332:732–5, May 2011.

[177] E F Petricoin, J L Hackett, L J Lesko, R K Puri, S I Gutman, K Chumakov, J Woodcock, D W Feigal, K C Zoon, and F D Sistare. Medical applications of microarray technologies: a regulatory science perspective. *Nature Genetics*, 32:474–9, December 2002.

[178] R J Prill, D Marbach, J Saez-Rodriguez, P K Sorger, L G Alexopoulos, X Xue, N D Clarke, G Altan-Bonnet, and G Stolovitzky. Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PloS one*, 5(2):e9202, January 2010.

[179] R J Prill, J Saez-Rodriguez, L G Alexopoulos, P K Sorger, and G Stolovitzky. Crowdsourcing network inference: the DREAM predictive signaling network challenge. *Science Signaling*, 4(189):mr7, September 2011.

[180] A Raj, C S Peskin, D Tranchina, D Y Vargas, and S Tyagi. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, 4(10):e309, October 2006.

[181] A Raj and A van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26, October 2008.

[182] J M Raser and E K O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304:1811–4, June 2004.

[183] M Rathinam, L R Petzold, Y Cao, and D T Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *The Journal of Chemical Physics*, 119(24):12784, 2003.

[184] K J Regehr, M Domenech, J T Koepsel, K C Carver, S J Ellison-Zelski, W L Murphy, L A Schuler, E T Alarid, and D J Beebe. Biological implications of polydimethylsiloxane-based microfluidic cell culture. *Lab on a Chip*, 9(15):2132–9, August 2009.

[185] S Regot, J J Hughey, B T Bajar, S Carrasco, and M W Covert. High-sensitivity measurements of multiple kinase activities in live single cells. *Cell*, 157(7):1724–1734, June 2014.

[186] R J D Reid, M Lisby, and R Rothstein. Cloning-free genome alterations in saccharomyces cereuisiae using adaptamer-mediated PCR. *Methods in Enzymology*, 350:258–77, 2002.

[187] R Rinott, A Jaimovich, and N Friedman. Exploring transcription regulation through cell-to-cell variability. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15):6329–34, April 2011.

[188] G O Roberts and J S Rosenthal. Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability*, 16(4):2123–39, November 2006.

[189] J Ruess, A Milias-Argeitis, and J Lygeros. Designing experiments to understand the variability in biochemical reaction networks. *Journal of The Royal Society Interface*, 10, 2013.

[190] J Ruess, A Milias-Argeitis, S Summers, and J Lygeros. Moment estimation for chemically reacting systems by extended Kalman filtering. *The Journal of Chemical Physics*, 135:165102, October 2011.

[191] E K Sackmann, A L Fulton, and D J Beebe. The present and future role of microfluidics in biomedical research. *Nature*, 507(7491):181–9, March 2014.

[192] I Sadowski, J Ma, S Triezenberg, and M Ptashne. GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, 335:563–4, 1988.

[193] H Saito and F Posas. Response to hyperosmotic stress. *Genetics*, 192(2):289–318, October 2012.

[194] M S Samoilov and A P Arkin. Deviant effects in molecular reaction pathways. *Nature Biotechnology*, 24(10):1235–40, October 2006.

[195] W Sandmann. Simultaneous stochastic simulation of multiple perturbations in biological network models. In *Computational Methods in Systems Biology*, pages 15–31. Springer, 2007.

[196] K A Schandel and D D Jenness. Direct evidence for ligand-induced internalization of the Yeast alpha-factor pheromone receptor. 14(11), 1994.

[197] J Schindelin, I Arganda-Carreras, E Frise, V Kaynig, M Longair, T Pietzsch, S Preibisch, C Rueden, S Saalfeld, B Schmid, J-Y Tinevez, D J White, V Hartenstein, K Eliceiri, P Tomancak, and A Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–82, July 2012.

[198] B L Schneider, W Seufert, B Steiner, Q H Yang, and A B Futcher. Use of polymerase chain reaction epitope tagging for protein tagging in Saccharomyces cerevisiae. *Yeast*, 11:1265–74, October 1995.

[199] J E Segall. Polarization of yeast cells in spatial gradients of alpha mating factor. *Proceedings of the National Academy of Sciences of the United States of America*, 90:8332–6, 1993.

[200] M Shaked and J G Shanthikumar. *Stochastic Orders (Springer Series in Statistics)*. Springer, 2007.

[201] N C Shaner, P A Steinbach, and R Y Tsien. A guide to choosing fluorescent proteins. *Nature Methods*, 2(12):905–9, December 2005.

[202] Ehud Shapiro, Tamir Biezuner, and Sten Linnarsson. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature reviews. Genetics*, 14(9):618–30, September 2013.

[203] S V Sharma, D Y Lee, B Li, M P Quinlan, F Takahashi, S Maheswaran, U McDermott, N Azizian, L Zou, M A Fischbach, K K Wong, K Brandstetter, B Wittner, S Ramaswamy, M Classon, and J Settleman. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1):69–80, April 2010.

[204] M A Sheff and K S Thorn. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast*, 21:661–70, June 2004.

[205] Y Shimada, M-P Gulli, and M Peter. Nuclear sequestration of the exchange factor Cdc24 by Far1 regulates cell polarity during yeast mating. *Nature Cell Biology*, 2, 2000.

[206] S Shimizu-Sato, E Huq, J M Tepperman, and P H Quail. A light-switchable gene promoter system. *Nature Biotechnology*, 20(10):1041–4, October 2002.

[207] R S Sikorski and P Hieter. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*, 122:19–27, 1989.

[208] D Silk, P D W Kirk, C P Barnes, T Toni, and M P H Stumpf. Model Selection in Systems Biology Depends on Experimental Design. *PLoS Computational Biology*, 10(6):e1003650, June 2014.

[209] B Snijder and L Pelkmans. Origins of regulated cell-to-cell variability. *Nature Reviews. Molecular Cell Biology*, 12:119–25, February 2011.

[210] R Spang. Diagnostic signatures from microarrays: a bioinformatics concept for personalized medicine. *Biosilico*, 1(2):64–8, May 2003.

[211] S L Spencer, S Gaudet, J G Albeck, J M Burke, and P K Sorger. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, 459(7245):428–32, May 2009.

[212] T M Squires and S R Quake. Microfluidics: Fluid physics at the nanoliter scale. *Reviews of Modern Physics*, 77, 2005.

[213] F Stagge, G Y Mitronova, V N Belov, C A Wurm, and S Jakobs. Snap-, CLIP- and Halo-tag labelling of budding yeast cells. *PloS one*, 8(10):e78745, January 2013.

[214] V Stathopoulos and M A Girolami. Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Philosophical Transactions of the Royal Society A*, 371(1984):20110541, 2013.

[215] F Steinke, M Seeger, and K Tsuda. Experimental design for efficient identification of gene regulatory networks using sparse Bayesian models. *BMC Systems Biology*, 1, January 2007.

[216] D J Stephens and V J Allan. Light microscopy techniques for live cell imaging. *Science*, 300(5616):82–6, April 2003.

[217] G Stolovitzky, D Monroe, and A Califano. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Annals of the New York Academy of Sciences*, 1115(914):1–22, December 2007.

[218] G Stolovitzky, R J Prill, and A Califano. Lessons from the DREAM2 challenges. *Annals of the New York Academy of Sciences*, 1158:159–95, March 2009.

[219] G Storvik. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50(2):281–9, 2002.

[220] D Strickland, Y Lin, E Wagner, C M Hope, J Zayner, C Antoniou, T R Sosnick, E L Weiss, and M Glotzer. TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nature methods*, 9(4):379–84, April 2012.

[221] D M Suter, N Molina, D Gatfield, K Schneider, U Schibler, and F Naef. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332:472–4, April 2011.

[222] Y Taniguchi, P J Choi, G E Li, H Chen, M Babu, J Hearn, A Emili, and X S Xie. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329:533–8, July 2010.

[223] A L Tarca, M Lauria, M Unger, E Bilal, S Boue, K K Dey, J Hoeng, H Koeppl, F Martin, P Meyer, P Nandy, R Norel, M Peitsch, J J Rice, R Romero, G Stolovitzky, M Talikka, Y Xiang, and C Zechner. Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. *Bioinformatics*, 29(22):2892–9, November 2013.

[224] S Tay, J J Hughey, T K Lee, T Lipniacki, S R Quake, and M W Covert. Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature*, 466(7303):267–71, July 2010.

[225] T Tian and K Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *The Journal of Chemical Physics*, 121(21):10356–64, December 2004.

[226] M W Toepke and D J Beebe. PDMS absorption of small molecules and consequences in microfluidic applications. *Lab on a Chip*, 6(12):1484–6, December 2006.

[227] J E Toettcher, D Gong, W A Lim, and O D Weiner. Light-based feedback for controlling intracellular signaling dynamics. *Nature Methods*, 2011.

[228] M Tsai, A Kita, J Leach, R Rounsevell, J N Huang, J Moake, R E Ware, D A Fletcher, and W A Lam. In vitro modeling of the microvascular occlusion and thrombosis that occur in hematologic diseases using microfluidic technology. *The Journal of clinical investigation*, 122(1):408–18, January 2012.

[229] R Y Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67:509–44, January 1998.

[230] J Uhlendorf, P Hersen, and G Batt. Towards real-time control of gene expression: in silico analysis. In *The 18th World Congress of the International Federation of Automatic Control*. IFAC, 2011.

[231] J Uhlendorf, A Miermont, T Delaveau, G Charvin, F Fages, S Bottani, G Batt, and P Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences of the United States of America*, (19), August 2012.

[232] M Unger, S S Lee, M Peter, and H Koeppl. Pulse Width Modulation of Liquid Flows: Towards Dynamic Control of Cell Microenvironments. In *15th International Conference on Miniaturized Systems for Chemistry and Life Sciences*, pages 1567–9, Seattle, Washington, USA, 2011. CBMS.

[233] M A Unger, H P Chou, T Thorsen, A Scherer, and S R Quake. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science*, 288(5463):113–6, April 2000.

[234] J Valdez-Taubas and H R B Pelham. Slow diffusion of proteins in the yeast plasma membrane allows polarity to be maintained by endocytic cycling. *Current Biology*, 13(18):1636–40, September 2003.

[235] L A L van de Pasch, A J Miles, W Nijenhuis, N A C H Brabers, D van Leenen, P Lijnzaad, M K Brown, J Ouellet, Y Barral, G J P L Kops, and F C P Holstege. Centromere binding and a conserved role in chromosome stability for SUMO-dependent ubiquitin ligases. *PloS one*, 8(6):e65628, January 2013.

[236] A Varshavsky. The N-end rule: Functions, mysteries, uses. *Proceedings of the National Academy of Sciences of the United States of America*, 93:12142–9, October 1996.

[237] A Varshavsky. The N-end rule pathway of protein degradation. *Genes to Cells*, 2(1):13–28, January 1997.

[238] S Wang and T Hazelrigg. Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature*, 369:400–3, 1994.

[239] Y Wang, H Senoo, H Sesaki, and M Iijima. Rho GTPases orient directional sensing in chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49):E4723–32, December 2013.

[240] R Wedlich-Soldner, S Altschuler, L Wu, and R Li. Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase. *Science*, 299(5610):1231–5, February 2003.

[241] R Wedlich-Soldner, S C Wai, T Schmidt, and R Li. Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling. *The Journal of Cell Biology*, 166(6):889–900, September 2004.

[242] J White and E Stelzer. Photobleaching GFP reveals protein dynamics inside live cells. *Trends in Cell Biology*, 9(2):61–5, February 1999.

[243] G Whitesides and A Stroock. Flexible methods for microfluidics. *Physics Today*, pages 42–48, 2001.

[244] G M Whitesides. The origins and the future of microfluidics. *Nature*, 442(7101):368–73, July 2006.

[245] D J Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC, 1st edition, 2006.

[246] D J Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews. Genetics*, 10(2):122–33, February 2009.

[247] D J Wilkinson. Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology. *Bayesian Statistics*, 9(Wilkinson 2009):679–705, 2010.

[248] T T Wu, Y F Chen, T Hastie, E Sobel, and K Lange. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6):714–21, March 2009.

[249] Y Xia and G M Whitesides. Soft Lithography. *Annual Reviews Material Sciences*, 28(12):153–84, 1998.

[250] Z Xiao, N Zhang, D B Murphy, and P N Devreotes. Dynamic distribution of chemoattractant receptors in living cells during chemotaxis and persistent stimulation. *The Journal of Cell Biology*, 139(2):365–74, October 1997.

[251] P Yager, T Edwards, E Fu, K Helton, K Nelson, M R Tam, and B H Weigl. Microfluidic diagnostic technologies for global public health. *Nature*, 442(7101):412–8, July 2006.

[252] H Ye, M Daoud-El Baba, R W Peng, and M Fussenegger. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science*, 332:1565–8, June 2011.

[253] L Ye, M Zhang, L G Alexopoulos, P Sorger, and K F Jensen. Microfluidic devices for studying the response of adherent cells under short time stimuli treatment. In *11th International Conference on Miniaturized Systems for Chemistry and Life Sciences*, number October, pages 158–60, Paris, France, 2007. CBMS.

[254] J Yu, V A Smith, P P Wang, A J Hartemink, and E D Jarvis. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, 20(18):3594–603, December 2004.

[255] C Zechner, S Deb, and H Koeppl. Marginal dynamics of stochastic biochemical networks in random environments. In *European Control Conference*, pages 4269–74, 2013.

[256] C Zechner, P Nandy, M Unger, and H Koeppl. Optimal variational perturbations for the inference of stochastic reaction dynamics. In *51st IEEE Conference on Decision and Control*, pages 5336–41, Maui, Hawaii, USA, December 2012. IEEE.

[257] C Zechner, S Pelet, M Peter, and H Koeppl. Recursive Bayesian Estimation of Stochastic Rate Constants from Heterogeneous Cell Populations. In *IEEE Conference on Decision and Control and European Control Conference*, pages 5837–43. IEEE, December 2011.

[258] C Zechner, J Ruess, P Krenn, S Pelet, M Peter, J Lygeros, and H Koeppl. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21):8340–5, May 2012.

[259] C Zechner, M Unger, S Pelet, M Peter, and H Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*, 11(2):197–202, January 2014.

[260] D Zenklusen, D R Larson, and R H Singer. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural and Molecular Biology*, 15(12):1263–71, December 2008.

[261] I K Zervantonakis, S K Hughes-Alford, J L Charest, J S Condeelis, F B Gertler, and R D Kamm. Three-dimensional microfluidic model for tumor cell intravasation and endothelial barrier function. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34):13515–20, August 2012.

[262] Z Zi, W Liebermeister, and E Klipp. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in Saccharomyces cerevisiae. *PLoS one*, 5(3):e9522, January 2010.

CURRICULUM VITÆ

MICHAEL PETER UNGER
Born January 12th, 1985. Citizen of Austria.

2011 - 2014	**Doctorate, ETH Zurich, Switzerland** Automatic Control Laboratory, D-ITET Institute of Biochemistry, D-BIOL (affiliated) Member of the PhD program: *Systems Biology of Complex Diseases*, Life Science Zurich Graduate School (Dr. sc. ETH Zürich)
2009 - 2010	**Master Studies in Telematics,** **Graz University of Technology, Austria** (MSc.)
2005 - 2009	**Bachelor Studies in Telematics,** **Graz University of Technology, Austria** (BSc.)
1999 - 2004	**Higher Institute of Technical Education,** **Pinkafeld, Austria** (Matura)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

Final Version as of January 16, 2015 (version 1.1).